

PCTWORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : A01K 67/00, A61K 38/46, 39/395, C07H 21/00, C07K 16/18, 16/40, C12N 9/16, 15/12, 15/55, 15/63, C12Q 1/68	A1	(11) International Publication Number: WO 99/04626 (43) International Publication Date: 4 February 1999 (04.02.99)
(21) International Application Number: PCT/US98/15828 (22) International Filing Date: 28 July 1998 (28.07.98) (30) Priority Data: 60/053,936 28 July 1997 (28.07.97) US (71) Applicant (for all designated States except US): FOX CHASE CANCER CENTER [US/US]; 7701 Burholme Avenue, Philadelphia, PA 19111 (US). (72) Inventor; and (75) Inventor/Applicant (for US only): BELLACOSA, Alfonso [IT/US]; 516 Kendrick Street, Philadelphia, PA 19111 (US). (74) Agents: RIGAUT, Kathleen, D. et al.; Dann, Dorfman, Herrell and Skillman, Suite 720, 1601 Market Street, Philadelphia, PA 19103-2307 (US).		(81) Designated States: AU, CA, JP, US, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE). Published <i>With international search report.</i>
(54) Title: NOVEL GENE ENCODING A DNA REPAIR ENDONUCLEASE AND METHODS OF USE THEREOF (57) Abstract An isolated nucleic acid molecule encoding a human endonuclease, MED1, is disclosed. Like other mismatch repair genes which are mutated in certain cancers, MED1, encoding nucleic acids, proteins and antibodies thereto may be used to advantage in genetic or cancer screening assays. MED1, which recognizes and cleaves DNA, may also be used for the diagnostic detection of mutations and genetic variants.		

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

**NOVEL GENE ENCODING A DNA REPAIR ENDONUCLEASE
AND METHODS OF USE THEREOF**

FIELD OF THE INVENTION

This invention relates to the field of DNA repair. Specifically, a novel human gene and its encoded endonuclease are disclosed. The gene may be used beneficially as a marker for genetic screening, mutational analysis and for assessing drug resistance in transformed cells.

5 BACKGROUND OF THE INVENTION

Several publications are referenced in this application in order to more fully describe the state of the art to which this invention pertains. The disclosure of each of these publications is incorporated
10 by reference herein.

Mismatch repair stabilizes the cellular genome by correcting DNA replication errors and by blocking recombination events between divergent DNA sequences. The mechanism responsible for strand-specific correction
15 of mispaired bases has been highly conserved during evolution. Eukaryotic homologs of bacterial Muts and MutL, which are believed to play key roles in mismatch repair recognition and initiation of repair, have been identified in yeast and mammalian cells. Inactivation
20 of genes encoding these activities results in large increases in spontaneous mutability, and in the case of humans and rodents, predisposition to tumor development.

Lynch syndrome or hereditary nonpolyposis colon cancer (HNPCC) is an autosomal dominant disease, which
25 accounts for approximately 1-5% of all colorectal cancer cases. In this syndrome, colorectal tumors are frequently associated with extracolonic malignancies, such as cancers of the endometrium, stomach, ovary, brain, skin and urinary tract. Tumors from HNPCC
30 patients harbor a genome-wide DNA replication/repair defect. Due to the lack of pathognomonic morphological

or biomolecular markers, HNPCC has traditionally posed unique problems to clinicians and geneticists alike, both in terms of diagnosis and clinical management.

Recent breakthroughs in molecular biology have partially elucidated the pathogenic mechanism of this syndrome. Germline mutations in any one of five genes encoding proteins that participate in a specialized DNA mismatch repair system give rise to a predisposition for cancer development in HNPCC families. Patients affected by HNPCC carry these mutations in genes which are involved in DNA mismatch repair. The DNA mismatch repair mechanism contributes to mutational avoidance and genetic stability, thus performing a tumor suppressor function. Loss or inactivation of the wild type allele in somatic cells leads to a dramatic increase of the spontaneous mutation rate. This, in turn, results in the accumulation of mutations in other tumor suppressor genes and oncogenes, ultimately leading to neoplastic transformation.

Microsatellites are repeating sequences that are distributed throughout the human genome, most commonly (A)_n/(T)_n and (CA)_n/(GT)_n. Their function is unknown, but they are useful in genetic linkage studies because of their high degree of polymorphism and normally stable inheritance. Several of the genes responsible for HNPCC have been identified using analysis of mutation rate in DNA microsatellites. Mutations of mismatch repair genes can be detected in a subset of sporadic colonic and extracolonic cancers which exhibit variability in the length of microsatellite sequences. This variability is often referred to as microsatellite instability.

Investigators in the field (Peltomaki et al., (1993) *Science* 260:810-812) have discovered that most colorectal cancers from HNPCC patients show microsatellite instability. These studies revealed that the length of microsatellite DNA at different loci varies between tumor DNA and non-tumor DNA from the same

patient. The phrase "replication error positive" (RER+) has been used to describe such tumors. It should be noted that only about 70% of HNPCC cases and only about 65% of sporadic tumors with microsatellite instability carry mutations in the known mismatch repair genes (hMSH2, hMLH1, hPMS2, hMSH6 and hPMS1) (Liu et al., (1996) *Nature Medicine* 2:169-174). The remaining 30-35% of the cases have an as yet unidentified mismatch repair genetic defect. Thus, there is a pressing need to identify the other active components in the DNA mismatch repair pathway, as mutations in these genes may result in an increased propensity for cancer.

The Fragile X or Martin Bell syndrome is the most common single recognized form of inherited mental retardation. Fifty percent of all X-linked mental retardation may be attributable to the Fragile X syndrome. The disorder is found in all ethnic groupings with a frequency of 0.3-1 per 1000 males and 0.2-0.6 per 1000 females. The full clinical syndrome, which is found in approximately 60% of affected males, consists of moderate mental retardation with an IQ typically in the range 35-50, elongated facies with large everted ears, and macroorchidism. This syndrome is unusual in that it is associated with the appearance of a fragile site on the long arm of the X chromosome at Xq27.3 (Sutherland, G.R., (1977) *Science* 197:256-266). This can be visualized cytogenetically in metaphase chromosomes prepared from lymphocytes of affected individuals which have been cultured under conditions of folate deficiency or thymidine stress. The study of the segregation of polymorphic markers within fragile X families has confirmed that the mutation lies in the same region of the X-chromosome as that exhibiting cytogenetic fragility.

There is an imbalance of penetrance of the phenotype associated with this syndrome in the different generations of kindreds in which the mutation is

segregating. The likelihood of developing mental impairment depends on an individual's position in the pedigree. As the mutation progresses through the generations, the risk of mental impairment increases.

5 These observations are not consistent with classical X linkage and are collectively known as the Sherman paradox. Hypotheses based on these observations have suggested that the mutation exists in two forms- a premutation and a full mutation form. Nonpenetrant
10 individuals are said to carry a premutation chromosome, that is a chromosome which has no abnormal phenotypic effect but which is capable of progressing to a fully penetrant mutation on passage through a female oogenesis.

15 Two alterations in the DNA at the fragile X site have been identified: abnormal amplification of a CpG-rich DNA sequence (a CpG island) and hypermethylation of such sequences. The molecular basis of the amplification is the expansion of a CGG triplet
20 microsatellite into large arrays. In individuals expressing the full clinical phenotype, the DNA in this region becomes hypermethylated, leading to the transcriptional shut down of the gene FMR-1 (fragile X mental retardation 1) which is transcribed across this
25 region. It is the loss of gene expression that is thought to account for the clinical phenotype. It has been postulated that in Fragile X syndrome, expansion of the (CGG)_n repeat from premutation to full mutation may be related to an aberrant (misdirected) DNA mismatch
30 repair event. This may be favored by the transient lack of multiple methyl signals in the CGG repeat as well as in flanking single copy sequences during early stages of embryonal development. Similar to Fragile X syndrome, defective DNA mismatch repair may play a role in the
35 expansion of triplet repeats associated with several disorders such as myotonic dystrophy, Huntington's disease, spino-cerebellar ataxias and Kennedy's disease.

The isolation of nucleic acids and proteins which when mutated give rise to these various disorders enables the development of diagnostic and prognostic kits for assessing patients at risk. The biochemical
5 characterization of the genes encoding the components of the DNA mismatch repair system may ultimately facilitate gene replacement therapies for use in the treatment of malignancy and other inherited genetic disorders.

10 SUMMARY OF THE INVENTION

This invention provides novel, biological molecules useful for identification, detection, and/or regulation of components in the complex DNA recognition/repair pathway. According to one aspect of the invention, an
15 isolated nucleic acid molecule is provided which includes a sequence encoding an endonuclease protein of a size between about 60 and 75 kilodaltons. The encoded protein, referred to herein as MED1 (methyl-CpG binding endonuclease 1) comprises a tripartite structure
20 including an amino terminal methyl-CpG binding domain with significant homology to the rat protein, MeCP2 and the human protein, PCM1, a central region rich in positively-charged amino acids which contains nuclear localization signals, and a carboxy terminal catalytic
25 domain which shares homology with several bacterial endonucleases involved in DNA repair. The protein demonstrates significant binding affinity for hMLH1 and mMLH2. In a preferred embodiment of the invention, an isolated nucleic acid molecule is provided that includes
30 a cDNA encoding a human endonuclease protein MED1. In a particularly preferred embodiment, the human endonuclease protein has an amino acid sequence the same as Sequence I.D. No. 2. An exemplary nucleic acid molecule of the invention comprises Sequence I.D. No. 1.

35 According to another aspect of the present invention, an isolated nucleic acid molecule is provided, which has a sequence selected from the group

consisting of: (1) Sequence I.D. No. 1; (2) a sequence specifically hybridizing with preselected portions or all of the complementary strand of Sequence I.D. No. 1; a sequence encoding preselected portions of Sequence I.D. No. 1, (3) a sequence encoding part or all of a polypeptide having amino acid Sequence I.D. No. 2. Such partial sequences are useful as probes to identify and isolate homologues of the endonuclease gene of the invention. Accordingly, isolated nucleic acid sequences encoding natural allelic variants of Sequence I.D. No. 1 are also contemplated to be within the scope of the present invention. The term natural allelic variants will be defined hereinbelow.

In yet another embodiment of the invention, isolated genomic DNA molecules are provided which encode the Med-1 protein of the invention. These nucleic acids (SEQ ID NO: 21 and 22) may be used to advantage in screening assays which identify germline and somatic mutations in the DNA encoding Med-1.

The present invention also provides MED1 genomic nucleic acid of mouse or human origin having a sequence substantially the same as that contained in phage stocks as deposited on 28 July 1998 at the American Type Culture Collection, 10801 University Blvd, Manassas, Virginia 20110-2209 USA, under the terms of the Budapest Treaty with accession number: not yet assigned.

MED1 polypeptide may conveniently be obtained by introducing expression vectors into host cells in which the vector is functional, culturing the host cells so that the MED1 polypeptide is produced and recovering the MED1 polypeptide from the host cells or the surrounding medium. Vectors comprising nucleic acid according to the present invention and host cells comprising such vectors or nucleic acid form further aspects of the present invention.

According to another aspect of the present invention, an isolated human endonuclease protein is

provided which has a deduced molecular weight of between about 60 kDa and 75 kDa. The protein comprises an amino-terminal methyl-CpG binding domain with significant homology to the rat protein MeCP2 and the human protein PCM1, a central region rich in positively-charged amino acids which contains nuclear localization signals, and a carboxy terminal catalytic domain which shares homology with several bacterial endonucleases involved in DNA repair. In a preferred embodiment of the invention, the protein is of human origin, and has an amino acid sequence the same as Sequence I.D. No. 2. In a further embodiment the protein may be encoded by natural allelic variants of Sequence I.D. No. 1. Inasmuch as certain amino acid variations may be present in a MED1 protein encoded by a natural allelic variant, such proteins are also contemplated to be within the scope of the invention.

According to another aspect of the present invention, antibodies immunologically specific for the proteins described hereinabove are provided.

Various terms relating to the biological molecules of the present invention are used hereinabove and also throughout the specifications and claims. The terms "specifically hybridizing," "percent similarity" and "percent identity (identical)" are defined in detail in the description set forth below.

With reference to nucleic acids of the invention, the term "isolated nucleic acid" is sometimes used. This term, when applied to DNA, refers to a DNA molecule that is separated from sequences with which it is immediately contiguous (in the 5' and 3' directions) in the naturally occurring genome of the organism from which it originates. For example, the "isolated nucleic acid" may comprise a DNA or cDNA molecule inserted into a vector, such as a plasmid or virus vector, or integrated into the DNA of a prokaryote or eukaryote.

With respect to RNA molecules of the

invention, the term "isolated nucleic acid" primarily refers to an RNA molecule encoded by an isolated DNA molecule as defined above. Alternatively, the term may refer to an RNA molecule that has been sufficiently separated from RNA molecules with which it would be associated in its natural state (i.e., in cells or tissues), such that it exists in a "substantially pure" form (the term "substantially pure" is defined below).

With respect to protein, the term "isolated protein" or "isolated and purified protein" is sometimes used herein. This term refers primarily to a protein produced by expression of an isolated nucleic acid molecule of the invention. Alternatively, this term may refer to a protein which has been sufficiently separated from other proteins with which it would naturally be associated, so as to exist in "substantially pure" form.

The term "substantially pure" refers to a preparation comprising at least 50-60% by weight the compound of interest (e.g., nucleic acid, oligonucleotide, protein, etc.). More preferably, the preparation comprises at least 75% by weight, and most preferably 90-99% by weight, the compound of interest. Purity is measured by methods appropriate for the compound of interest (e.g. chromatographic methods, agarose or polyacrylamide gel electrophoresis, HPLC analysis, and the like).

With respect to antibodies of the invention, the term "immunologically specific" refers to antibodies that bind to one or more epitopes of a protein of interest (e.g., MED1), but which do not substantially recognize and bind other molecules in a sample containing a mixed population of antigenic biological molecules.

With respect to oligonucleotides, the term "specifically hybridizing" refers to the association between two single-stranded nucleotide molecules of sufficiently complementary sequence to permit such

hybridization under pre-determined conditions generally used in the art (sometimes termed "substantially complementary"). In particular, the term refers to hybridization of an oligonucleotide with a substantially complementary sequence contained within a single-stranded DNA or RNA molecule of the invention, to the substantial exclusion of hybridization of the oligonucleotide with single-stranded nucleic acids of non-complementary sequence.

10 The present invention also includes active portions, fragments, derivatives and functional mimetics of the MED1 polypeptide or protein of the invention.

An "active portion" of MED1 polypeptide means a peptide which is less than said full length MED1 polypeptide, but which retains its essential biological activity, e.g., methyl-CpG DNA binding and/or endonuclease activity.

20 A "fragment" of the MED1 polypeptide means a stretch of amino acid residues of at least about five to seven contiguous amino acids, often at least about seven to nine contiguous amino acids, typically at least about nine to thirteen contiguous amino acids and, most preferably, at least about twenty to thirty or more contiguous amino acids. Fragments of the MED1 polypeptide sequence, antigenic determinants or epitopes are useful for raising antibodies to a portion of the MED1 amino acid sequence.

25 A "derivative" of the MED1 polypeptide or a fragment thereof means a polypeptide modified by varying the amino acid sequence of the protein, e.g. by manipulation of the nucleic acid encoding the protein or by altering the protein itself. Such derivatives of the natural amino acid sequence may involve insertion, addition, deletion or substitution of one or more amino acids, without fundamentally altering the essential activity of the wildtype MED1 polypeptide.

35 "Functional mimetic" means a substance which may not

contain an active portion of the MED1 amino acid sequence, and probably is not a peptide at all, but which retains the essential biological activity of natural MED1 polypeptide.

5 The nucleic acids, proteins/polypeptides, peptides and antibodies of the present invention may be used to advantage as markers for diagnosis and prognosis of those at risk for colon and other cancers. The molecules may also be useful in the diagnosis and/or
10 treatment of Fragile X syndrome and other diseases characterized by triplet repeat expansion. The MED1 molecules of the invention may also be used as research tools and will facilitate the elucidation of the mechanistic action of the novel genetic and protein
15 interactions involved in the maintenance of DNA fidelity.

Thus, the present invention also provides nucleic acid molecules, polypeptides and/or antibodies as mentioned above for use in medical treatment.

20 Further, the present invention provides use of a nucleic acid molecule, polypeptide and/or antibody in the preparation of a medicament for treating cancer, in particular, colorectal cancer.

In a further aspect of the present invention, there
25 is provided a kit for detecting mutations in the MED1 gene associated with cancer, or a susceptibility to cancer, the kit comprising one or more nucleic acid probes capable of binding and/or detecting a mutated MED1 nucleic acid. Alternatively, the kit may comprise
30 one or more antibodies capable of specifically binding and/or detecting a mutated MED1 nucleic acid or amino acid sequence or a pair of oligonucleotide primers having sequences corresponding to, or complementary to a portion of the nucleic acid sequence set out in
35 Sequence I. D. No. 1 or 5 for use in amplifying a MED1 nucleic acid sequence or mutant allele thereof.

In yet another aspect of the invention, transgenic

animals are provided which in growth and development are useful for elucidating the role of MED1. Isolation of the mouse genomic DNA also facilitates the production of MED1 knock-out mice.

5 Aspects and embodiments of the present invention will now be illustrated, by way of example, with reference to the accompanying figures. Further aspects and embodiments will be apparent to those skilled in the art. All documents mentioned in this text are
10 incorporated herein by reference.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 depicts EGY191 yeast cells cotransformed with a combination of plasmids as indicated in the
15 figure along with pSH18-34. The yeast so transformed were then selected on uracil-minus, histidine-minus tryptophan-minus glucose yeast medium to select for the presence of all plasmids. Individual transformants were replated either onto uracil-minus, histidine-minus,
20 tryptophan-minus, leucine-minus galactose yeast medium to score activation of the LEU2 reporters (left panel) or onto uracil-minus, histidine-minus, tryptophan-minus galactose yeast medium containing 5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside (X-gal) to score
25 activation of the LacZ reporters (right panel). Growth on leucine-minus plates and blue-color formation on LEX-gal plates illustrate the specificity of the interaction between f5/MED1 and hMLH1. All interactions were galactose specific. The interaction shown between K-
30 rev-1 and Krit1 represents a positive control.

Figure 2 depicts a Northern blot showing the localization of MED1 mRNA in all tested tissues. A 2.4 kb transcript is observed and high levels of mRNA
35 expression is detected in heart, skeletal muscle and pancreas. The size of the molecular weight standards is indicated in kb.

Figure 3 shows an alignment of the cDNA of Sequence I.D. No. 1 and its encoded endonuclease protein, Sequence I.D. No. 2.

5 Figure 4A depicts homology analysis of the deduced amino acid sequence of MED1 and several other endonucleases involved in DNA recognition and repair. Figure 4B depicts homology analysis of the deduced amino acid sequence of MED1 and the methyl-CpG binding domain
10 of the rat protein, MeCP2. Figure 4C depicts homology analysis of the deduced amino acid sequence of MED1 and the methyl-CpG binding domain of the human protein, PCM1.

15 Figure 5 is a schematic diagram illustrating the domain organization of MED1 protein. The methyl-CpG binding domain (MBD) and the endonuclease domain (endo) are highlighted. Numbers indicate amino acid position. The bar below the schematic diagram indicates the
20 portion of the protein encoded by the original f5 clone.

Figure 6 is an autoradiograph showing the results of coupled in vitro transcription and translation of the MED1 open reading frame. Two polypeptides of 70 and 65
25 kD are synthesized by pcDNA3-MED1 constructs. In control reactions, lacking the MED1 cDNA, these polypeptides are not synthesized.

Figures 7A and 7B show a schematic diagram (Fig. 7A) of carboxy- and amino-terminal hemagglutinin-tagged (HT) MED1 proteins and a Western Blot (Fig. 7B) showing
30 protein expression following transfection of the constructs into NIH 3T3 cells. A band of approximately 72 kD is present in cells transfected with the carboxyterminally tagged MED1-HT. This band co-migrates
35 with the one present in HT-MED1-M1 transfectants, indicating that the first ATG at nucleotide position 142

is the initiation codon in vivo.

Figure 8 is a partial metaphase spread of human chromosomes showing the chromosomal localization of MED1 by FISH. Hybridization is detected on chromosome 3q21 (arrow). An elongated chromosome 3 is shown in the inset.

Figures 9A and 9B are gels and blots demonstrating the nuclease activity of the recombinant endonuclease domain. Figure 9A is a Coomassie-stained SDS-PAGE showing IPTG induction of the bacterially-expressed 18-22-kD MED1 endonuclease domain (codons 455-580) (arrowhead, left panel). In a parallel SDS-PAGE nuclease activity gel (containing heat-denatured calf thymus DNA), the IPTG-induced 18-22-kD MED1 endonuclease domain is negatively stained with the DNA dye, toluidine blue (arrowhead, right panel). P, pellet of 10,000x g centrifugation; S, supernatant of 10,000x g centrifugation. Figure 9B shows endonuclease activity of recombinant wild-type MED1. The entire wild-type MED1 and a deletion mutant lacking the endonuclease domain (Δ endo) were expressed in bacteria, purified by nickel-agarose chromatography and stained with Coomassie following SDS-PAGE (left panel). Increasing amounts of the wild-type and Δ endo mutant (22 to 175 ng) were incubated with 500 ng of the 3.9 kb supercoiled plasmid pCR2 (Invitrogen) at 37°C for 30. Reaction products were separated on a 1% agarose gel buffered in 1x TAE and containing 0.25 μ g/ml ethidium bromide (right panel). Wild-type MED1, but not Δ endo, generated nicked and linearized DNA. M, lambda/HindIII digest size standards; I, input plasmid DNA, incubated with reaction buffer only.

35

Figure 10A is an autoradiograph showing the results of a mobility shift assay of 293 cell lysates expressing

the fusion protein Flag-MED1/f5. Flag-peptide eluates from anti-Flag immunoprecipitations of Flag-MED1/f5-expressing 293 cells demonstrate binding activity when incubated with a ^{32}P -labeled double-stranded oligonucleotide containing five fully methylated CpG sites. A mobility shift assay of recombinant MED1 MBD (codons 1-154) with methylated and unmethylated DNA probes is shown in Figure 10B. The purified MED1 MBD demonstrates binding activity when incubated with a ^{32}P -labeled double-stranded oligonucleotide containing five methylated CpG sites (lane 2). Binding is abolished by pre-incubation with a 100-fold excess of the cold methylated oligonucleotide (lane 3), but not of the cold unmethylated oligonucleotide (lane 4). No binding is detected when the unmethylated probe is used (lanes 5-8)

Figures 11A and 11B are autoradiographs showing the coimmunoprecipitation of hMSH2 with Flag-MED1/f5. Fig. 11A shows a band reacting with the anti-hMSH2 antibody. Comigration with hMSH2 is detected by western blotting in anti-FLAG immunoprecipitates from Flag-MED1/f5 transfected cells but not control cells. Fig. 11B is a western blot of a parallel gel with the anti-FLAG antibody confirming expression of the Flag-MED1/f5 construct in transfected 293 cells.

Co-immunoprecipitation of MED1 and MLH1 from human cells is shown in Figure 11C. A band reacting with the anti-MLH1 antibody and comigrating with MLH1 is detected by western blotting in anti-hemagglutinin immunoprecipitates from HT-MED1/CMV5-transfected HEK-293 cells and not from CMV5-transfected control cells (upper panel). Western blotting of a parallel gel with the anti-hemagglutinin antibody confirms expression of the HT-MED1 construct in transfected HEK-293 cells (lower panel). Lysis buffers contained 0.5% NP-40 (lanes 1-4), 0.2% NP-40 (lanes 5-6) or 1% Triton X-100 (lanes 7-8).

Figure 12 is a schematic diagram depicting a model for strand targeting in eukaryotic mismatch repair. Recognition of the hemimethylated d(GATC) site by *E. coli* MthH (upper panel) is paralleled by recognition of the hemimethylated CpG site by human MED1 (lower panel).

Figure 13 shows a series of MED1 mutations which have been isolated from colon cancer patients. Figures 13A and 13B show MED1 sequencing electropherograms (ABI) of three colon tumor DNAs and a normal control DNA. Tumors c220T and c226T harbor an apparently heterozygous adenine deletion at the (A)10 track (codons 310-313) with predicted frameshift and stop at codon 317 (Fig. 13A). The same mutation was also found in tumor c18T. Tumor c215T harbors an apparently heterozygous adenine deletion at the (A)6 track (codons 280-282) with predicted frameshift and stop at codon 302 (Fig. 13B). Figure 13C shows a schematic diagram of the truncated products predicted to be encoded by the mutant MED1 alleles in the indicated tumors.

Figure 14 is a schematic diagram of the genomic structure of the human MED1 gene (lambda clone MED1 HGL #16). The position of the eight exons is indicated. Numbers above the exon boxes refer to exon number; numbers below the exon boxes refer to the size of the exons in base pairs. Exon 1 and part of the intervening intron between exon 1 and exon 2 was cloned by PCR (indicated by the hatched line). The start (ATG) and stop (TAA) codons are marked. E: restriction site for the enzyme EcoRI.

Figure 15 is a blot showing the conservation of the MED1 gene ("Zooblot"). A low stringency Southern blot of genomic DNA from indicated vertebrate species reveals bands cross-hybridizing with a human MED1 cDNA probe in mammals (panel A) and non-mammalian vertebrates (panel

B). The migration and size (in kilobase pairs) of the DNA standards are indicated.

Figure 16 shows a schematic of the genomic structure of the mouse MED1 gene (lambda clone MED1 MGL #3). The position of seven exons is indicated. Numbers above the exon boxes refer to exon number; numbers below the exon boxes refer to the size of the exons in base pairs. The size and position of the exon 1 are not well defined (as indicated by the dotted line). The start (ATG) codon is marked. The stop codon is presumably located in exon 8 which is not contained in this lambda clone. E: restriction site for the enzyme EcoRI; S: restriction site for the enzyme Sall.

Figure 17 shows the nucleotide sequence (SEQ ID NO: 5) of the mouse cDNA MED1 sequence assembled by juxtaposition of seven exons derived from the genomic clone MED1 MGL #3.

Figure 18 shows a comparison of the predicted mouse MED1 protein sequence with the human MED1 protein sequence. Upper sequence: mouse MED1; lower sequence: human MED1. Identical amino acids between the two sequences are indicated by a line, similar amino acids by one (low similarity) or two dots (high similarity).

Figure 19 shows the intron and exon sequences of the mouse genomic clone encoding MED1. Exon sequences are shown in upper case; intron sequences are shown in lower case. The splice donor (gt) and acceptor (ga) sites are in bold.

Figure 20 shows the intron and exon sequences of the human genomic clone encoding MED1. Exon sequences are shown in upper case; intron sequences are shown in lower case. The splice donor (gt) and acceptor (ga)

sites are in bold.

DETAILED DESCRIPTION OF THE INVENTION

Hereditary Non-Polyposis Colorectal Cancer (HNPCC),
5 or Lynch Syndrome, is an autosomal dominant disorder characterized by early onset colorectal tumors. As noted above, tumors from HNPCC patients harbor a genome-wide DNA replication/repair defect, the hallmark of which is length instability of microsatellite repeat
10 sequences. Patients affected by HNPCC carry a germline mutation in genes involved in DNA mismatch repair, a specialized system which handles base-base mismatches, short insertions/deletions and recombination-derived heteroduplexes (Kolodner, R.D., (1995) *Trends in*
15 *Biochem. Sci.* 20:397-4053; Modrich and Lahue, (1996) *Annu. Rev. Biochem.* 65:101-133). The mismatch repair pathway contributes to mutational avoidance and genetic stability, thus performing a tumor suppressor function. Loss or inactivation of the wild type allele in somatic
20 cells leads to a dramatic increase of the spontaneous mutation rate. This, in turn, results in the accumulation of mutations in other tumor suppressor genes and oncogenes, ultimately leading to neoplastic transformation (Bellacosa et al., (1996) *Am. J. of Med.*
25 *Genetics* 62:353-364). Similarly to other genes involved in tumor suppression, mutations of mismatch repair genes can be detected in a subset of sporadic colonic and extracolonic cancers which exhibit microsatellite instability (Liu et al., 1996, supra).

30 Any one of five DNA mismatch repair genes (hMSH2, hMLH1, hPMS2, hMSH6 and hPMS1) is found to be mutated in the germline DNA of HNPCC patients (Liu et al., 1996, supra). These genes encode human homologues of the *E. coli* mismatch repair proteins MutS and MutL, which
35 belong to the methyl-directed mismatch repair system (Kolodner, R.D., 1995, supra). Repair by this system involves 10 biochemical activities and is organized in

3 sequential steps of initiation, excision and
resynthesis (Modrich, P., 1991) *Ann. Rev. Genet.* 25:229-
253). During initiation, the mismatch is detected and
5 a single-strand cut is made on the newly synthesized DNA
strand which contains the mutation. Then, single-strand
exonucleases (exo I, exo VII, RecJ) excise a span of
about 1-2 kbp containing the mismatch and finally
resynthesis by DNA polymerase III takes place. The
10 products of the mutSLH genes mediate the initiation
step. MutS detects and binds to the mismatch. Through
an interaction with MutL, which likely functions as an
interface with MutS, the single-strand endonuclease MutH
is activated and cuts the DNA strand carrying the
mutation (Modrich, P., 1991, *supra*).

15 A similar biochemical pathway has been identified
in eukaryotic cells, and it is also characterized by
strand-specificity and bidirectional excision capability
(Fang and Modrich, (1993) *J. Biol. Chem.* 268:11838-
11844). In the bacterial system, MutH has the pivotal
20 role of identifying the newly synthesized strand, i.e.
the strand carrying the mutation. Without this function
there would be a 50% chance of initiating repair on the
parental strand, thereby stabilizing the mutation. MutH
identifies and cleaves the new strand by virtue of its
25 transient lack of adenine methylation at d(GATC) sites
(Modrich, P., 1991, *supra*). Despite its crucial
function, homologues of MutH, i.e. eukaryotic mismatch
repair endonucleases, have not been identified to date.
Furthermore, the molecular determinants of strand
30 discrimination in eukaryotic cells - which lack d(GATC)
methylation - are not presently known (Kolodner, R.D.,
1995, *supra*; Modrich and Lahue, 1996, *supra*). In order
to gain insight into the mechanisms of strand
recognition, it is essential to identify the eukaryotic
35 functional homologue of the MutH endonuclease. Due to
its proposed central role in mismatch repair,
inactivation of this enzyme could be responsible for at

least some cases of HNPCC.

As mentioned previously, aberrant DNA methylation may also play a role in Fragile X Syndrome. After semi-conservative replication of DNA, the mismatch repair system is able to use the conserved strand as a template to correct mismatches resulting from replication errors which are by definition in the newly synthesized strand. DNA replication results in a transient state of hemimethylation in which methylation occurs only on the template strand. In Fragile X Syndrome, the CGG repeats and subsequent expansion of these repeats may be triggered by undermethylation leading to misdirection of DNA mismatch repair. MED1 encoded proteins may play a pivotal role in this aberrant DNA replication/repair event. As mentioned earlier, this could also be the case for other diseases associated with repeat expansion, such as myotonic dystrophy, Huntington's disease, spino-cerebellar ataxias and Kennedy's disease.

The genomic and cDNA cloning of MED1, the DNA molecule of the invention, which encodes a protein bearing homology to bacterial endonucleases is described in detail below. Analysis of the predicted amino acid sequence of the MED1 protein suggests a putative mechanism of strand recognition based on cytosine methylation at CpG sites. Like other DNA recognition and repair genes which are mutated in HNPCC as well as in sporadic cancers with microsatellite instability, MED1 is a candidate nucleic acid for cancer genetic testing, both in HNPCC families and in sporadic cancers with microsatellite instability. Aberrant MED1 activity may also be associated with Fragile X Syndrome and other diseases characterized by triplet repeat expansion.

I. Preparation of MED1-Encoding Nucleic Acid Molecules, MED1 Proteins, and Antibodies Thereto

A. Nucleic Acid Molecules

Nucleic acid molecules encoding the MED1 endonuclease of the invention may be prepared by two

general methods: (1) Synthesis from appropriate nucleotide triphosphates, or (2) Isolation from biological sources. Both methods utilize protocols well known in the art.

5 The availability of nucleotide sequence information, such as the nearly full length cDNA having Sequence I.D. No. 1, enables preparation of an isolated nucleic acid molecule of the invention by
10 oligonucleotide synthesis. Synthetic oligonucleotides may be prepared by the phosphoramidite method employed in the Applied Biosystems 38A DNA Synthesizer or similar devices. The resultant construct may be purified according to methods known in the art, such as high performance liquid chromatography (HPLC). Long, double-
15 stranded polynucleotides, such as a DNA molecule of the present invention, must be synthesized in stages, due to the size limitations inherent in current oligonucleotide synthetic methods. Thus, for example, a 2.4 kb double-
20 stranded molecule may be synthesized as several smaller segments of appropriate complementarity. Complementary segments thus produced may be annealed such that each segment possesses appropriate cohesive termini for attachment of an adjacent segment. Adjacent segments may be ligated by annealing cohesive termini in the
25 presence of DNA ligase to construct an entire 2.4 kb double-stranded molecule. A synthetic DNA molecule so constructed may then be cloned and amplified in an appropriate vector.

 Nucleic acid sequences encoding MED1 may be
30 isolated from appropriate biological sources using methods known in the art. In a preferred embodiment, a cDNA clone is isolated from a cDNA expression library of human origin. In an alternative embodiment, utilizing the sequence information provided by the cDNA sequence,
35 genomic clones encoding MED1 may be isolated. Alternatively, cDNA or genomic clones having homology with MED1 may be isolated from other species, such as

mouse, using oligonucleotide probes corresponding to predetermined sequences within the MED1 gene.

In accordance with the present invention, nucleic acids having the appropriate level of sequence homology with the protein coding region of Sequence I.D. No. 1 may be identified by using hybridization and washing conditions of appropriate stringency. For example, hybridizations may be performed, according to the method of Sambrook et al., (supra) using a hybridization solution comprising: 5X SSC, 5X Denhardt's reagent, 0.5-1.0% SDS, 100 µg/ml denatured, fragmented salmon sperm DNA, 0.05% sodium pyrophosphate and up to 50% formamide. Hybridization is carried out at 37-42°C for at least six hours. Following hybridization, filters are washed as follows: (1) 5 minutes at room temperature in 2X SSC and 0.5-1% SDS; (2) 15 minutes at room temperature in 2X SSC and 0.1% SDS; (3) 30 minutes-1 hour at 37°C in 1X SSC and 1% SDS; (4) 2 hours at 42-65°C in 1X SSC and 1% SDS, changing the solution every 30 minutes.

One common formula for calculating the stringency conditions required to achieve hybridization between nucleic acid molecules of a specified sequence homology is (Sambrook et al., 1989):

$$T_m = 81.5^{\circ}\text{C} + 16.6\text{Log} [\text{Na}^+] + 0.41(\% \text{ G+C}) - 0.63 (\% \text{ formamide}) - 600/\# \text{bp in duplex}$$

As an illustration of the above formula, using $[\text{Na}^+] = [0.368]$ and 50% formamide, with GC content of 42% and an average probe size of 200 bases, the T_m is 57°C. The T_m of a DNA duplex decreases by 1 - 1.5°C with every 1% decrease in homology. Thus, targets with greater than about 75% sequence identity would be observed using a hybridization temperature of 42°C. Such a sequence would be considered substantially homologous to the nucleic acid sequence of the present invention.

Nucleic acids of the present invention may be maintained as DNA in any convenient cloning vector. In

a preferred embodiment, clones are maintained in a plasmid cloning/expression vector, such as pBluescript (Stratagene, La Jolla, CA), which is propagated in a suitable *E. coli* host cell. Genomic clones of the invention encoding the human or mouse MED1 gene may be maintained in lambda phage FIX II (Stratagene).

MED1-encoding nucleic acid molecules of the invention include cDNA, genomic DNA, RNA, and fragments thereof which may be single- or double-stranded. Thus, this invention provides oligonucleotides (sense or antisense strands of DNA or RNA) having sequences capable of hybridizing with at least one sequence of a nucleic acid molecule of the present invention, such as selected segments of the cDNA having Sequence I.D. No. 1. Such oligonucleotides are useful as probes for detecting or isolating MED1 genes.

It will be appreciated by persons skilled in the art that variants (e.g., allelic variants) of these sequences exist in the human population, and must be taken into account when designing and/or utilizing oligos of the invention. Accordingly, it is within the scope of the present invention to encompass such variants, with respect to the MED1 sequences disclosed herein or the oligos targeted to specific locations on the respective genes or RNA transcripts. With respect to the inclusion of such variants, the term "natural allelic variants" is used herein to refer to various specific nucleotide sequences and variants thereof that would occur in a human population. Genetic polymorphisms giving rise to conservative or neutral amino acid substitutions in the encoded protein are examples of such variants. Additionally, the term "substantially complementary" refers to oligo sequences that may not be perfectly matched to a target sequence, but the mismatches do not materially affect the ability of the oligo to hybridize with its target sequence under the conditions described.

Thus, the coding sequence may be that shown in Sequence I.D. No. 1, or it may be a mutant, variant, derivative or allele of this sequence. The sequence may differ from that shown by a change which is one or more of addition, insertion, deletion and substitution of one or more nucleotides of the sequence shown. Changes to a nucleotide sequence may result in an amino acid change at the protein level, or not, as determined by the genetic code.

Thus, nucleic acid according to the present invention may include a sequence different from the sequence shown in Sequence I.D. No. 1 yet encode a polypeptide with the same amino acid sequence.

On the other hand, the encoded polypeptide may comprise an amino acid sequence which differs by one or more amino acid residues from the amino acid sequence shown in Sequence I.D. No. 2. Nucleic acid encoding a polypeptide which is an amino acid sequence mutant, variant, derivative or allele of the sequence shown in Sequence I.D. No. 2 is further provided by the present invention. Nucleic acid encoding such a polypeptide may show greater than 60% homology with the coding sequence shown in Sequence I.D. No. 1, greater than about 70% homology, greater than about 80% homology, greater than about 90% homology or greater than about 95% homology.

Also within the scope of the invention are antisense oligonucleotide sequences based on the MED1 nucleic acid sequences described herein. Antisense oligonucleotides may be designed to hybridize to the complementary sequence of nucleic acid, pre-mRNA or mature mRNA, interfering with the production of polypeptides encoded by a given DNA sequence (e.g. either native MED1 polypeptide or a mutant form thereof), so that its expression is reduced or prevented altogether. In addition to the MED1 coding sequence, antisense techniques can be used to target control sequences of the MED1 gene, e.g. in the 5' flanking

sequence of the MED1 coding sequence, whereby the antisense oligonucleotides can interfere with MED1 control sequences. The construction of antisense sequences and their use is described in Peyman and
5 Ulman, Chemical Reviews, 90:543-584, (1990), Crooke, Ann. Rev. Pharmacol. Toxicol., 32:329-376, (1992), and Zamecnik and Stephenson, Proc. Natl. Acad. Sci., 75:280-284, (1974).

The present invention provides a method of
10 obtaining nucleic acid of interest, the method including hybridization of a probe having part or all of the sequence shown in Sequence I.D. No. 1 or a complementary sequence, to target nucleic acid. Hybridization is generally followed by identification of successful
15 hybridization and isolation of nucleic acid which has hybridized to the probe, which may involve one or more steps of PCR.

Such oligonucleotide probes or primers, as well as the full-length sequence (and mutants, alleles,
20 variants, and derivatives) are useful in screening a test sample containing nucleic acid for the presence of alleles, mutants or variants, especially those that confer susceptibility or predisposition to cancers, the probes hybridizing with a target sequence from a sample
25 obtained from the individual being tested. The conditions of the hybridization can be controlled to minimize non-specific binding, and preferably stringent to moderately stringent hybridization conditions are used. The skilled person is readily able to design such
30 probes, label them and devise suitable conditions for hybridization reactions, assisted by textbooks such as Sambrook et al (1989) and Ausubel et al (1992).

In some preferred embodiments, oligonucleotides according to the present invention that are fragments of
35 the sequences shown in Sequence I.D. No. 1 or Sequence I.D. No. 5, or any allele associated with cancer susceptibility, are at least about 10 nucleotides in

length, more preferably at least 15 nucleotides in length, more preferably at least about 20 nucleotides in length. Such fragments themselves individually represent aspects of the present invention. Fragments and other
5 oligonucleotides may be used as primers or probes as discussed but may also be generated (e.g. by PCR) in methods concerned with determining the presence in a test sample of a sequence indicative of cancer susceptibility.

10 Methods involving use of nucleic acid in diagnostic and/or prognostic contexts, for instance in determining susceptibility to cancer, and other methods concerned with determining the presence of sequences indicative of cancer susceptibility are discussed below.

15 Nucleic acid according to the present invention may be used in methods of gene therapy, for instance in treatment of individuals with the aim of preventing or curing (wholly or partially) cancer. This too is discussed below.

20 **B. Proteins**

MED1 protein demonstrates methyl-CpG DNA binding and endonuclease activity. A full-length MED1 protein of the present invention may be prepared in a variety of ways, according to known methods. The
25 protein may be purified from appropriate sources, e.g., transformed bacterial or animal cultured cells or tissues, by immunoaffinity purification. However, this is not a preferred method due to the low amount of protein likely to be present in a given cell type at any
30 time. The availability of nucleic acid molecules encoding MED1 enables production of the protein using *in vitro* expression methods known in the art. For example, a cDNA or gene may be cloned into an appropriate *in vitro* transcription vector, such as pSP64 or pSP65 for
35 *in vitro* transcription, followed by cell-free translation in a suitable cell-free translation system, such as wheat germ or rabbit reticulocyte lysates. In

vitro transcription and translation systems are commercially available, e.g., from Promega Biotech, Madison, Wisconsin or BRL, Rockville, Maryland.

Alternatively, according to a preferred embodiment, larger quantities of MED1 may be produced by expression in a suitable prokaryotic or eukaryotic system. For example, part or all of a DNA molecule, such as the cDNA having Sequence I.D. No. 1, may be inserted into a plasmid vector adapted for expression in a bacterial cell, such as *E. coli*. Such vectors comprise the regulatory elements necessary for expression of the DNA in the host cell (e.g. *E. coli*) positioned in such a manner as to permit expression of the DNA in the host cell. Such regulatory elements required for expression include promoter sequences, transcription initiation sequences and, optionally, enhancer sequences.

The MED1 produced by gene expression in a recombinant prokaryotic or eukaryotic system may be purified according to methods known in the art. In a preferred embodiment, a commercially available expression/secretion system can be used, whereby the recombinant protein is expressed and thereafter secreted from the host cell, to be easily purified from the surrounding medium. If expression/secretion vectors are not used, an alternative approach involves purifying the recombinant protein by affinity separation, such as by immunological interaction with antibodies that bind specifically to the recombinant protein or nickel columns for isolation of recombinant proteins tagged with 6-8 histidine residues at their N-terminus or C-terminus. Alternative tags may comprise the FLAG epitope or the hemagglutinin epitope. Such methods are commonly used by skilled practitioners.

The MED1 proteins of the invention, prepared by the aforementioned methods, may be analyzed according to standard procedures. For example, such proteins may

be subjected to amino acid sequence analysis, according to known methods.

5 As discussed above, a convenient way of producing a polypeptide according to the present invention is to express nucleic acid encoding it, by use of the nucleic acid in an expression system. The use of expression systems has reached an advanced degree of sophistication today.

10 Accordingly, the present invention also encompasses a method of making a polypeptide (as disclosed), the method including expression from nucleic acid encoding the polypeptide (generally nucleic acid according to the invention). This may conveniently be achieved by growing a host cell in culture, containing such a vector, under
15 appropriate conditions which cause or allow production of the polypeptide. Polypeptides may also be produced in in vitro systems, such as reticulocyte lysate.

Polypeptides which are amino acid sequence variants, alleles, derivatives or mutants are also
20 provided by the present invention. A polypeptide which is a variant, allele, derivative, or mutant may have an amino acid sequence that differs from that given in Sequence I.D. No. 2 by one or more of addition, substitution, deletion and insertion of one or more
25 amino acids. Preferred such polypeptides have MED1 function, that is to say have one or more of the following properties: methyl-CpG DNA binding activity; endonuclease activity; immunological cross-reactivity with an antibody reactive with the polypeptide for which the sequence is given in Sequence I.D. No. 2; sharing an
30 epitope with the polypeptide for which the sequence is given in Sequence I.D. No. 2 (as determined for example by immunological cross-reactivity between the two polypeptides.

35 A polypeptide which is an amino acid sequence variant, allele, derivative or mutant of the amino acid sequence shown in Sequence I.D. No. 2 may comprise an

amino acid sequence which shares greater than about 35% sequence identity with the sequence shown, greater than about 40%, greater than about 50%, greater than about 60%, greater than about 70%, greater than about 80%, greater than about 90% or greater than about 95%. Particular amino acid sequence variants may differ from that shown in Sequence I.D. No.2 by insertion, addition, substitution or deletion of 1 amino acid, 2, 3, 4, 5-10, 10-20, 20-30, 30-40, 40-50, 50-100, 100-150, or more than 150 amino acids.

A polypeptide according to the present invention may be used in screening for molecules which affect or modulate its activity or function. Such molecules may be useful in a therapeutic (possibly including prophylactic) context.

The present invention also provides antibodies capable of immunospecifically binding to proteins of the invention. Polyclonal antibodies directed toward MED1 may be prepared according to standard methods. In a preferred embodiment, monoclonal antibodies are prepared, which react immunospecifically with various epitopes of MED1. Monoclonal antibodies may be prepared according to general methods of Köhler and Milstein, following standard protocols. Polyclonal or monoclonal antibodies that immunospecifically interact with MED1 can be utilized for identifying and purifying such proteins. For example, antibodies may be utilized for affinity separation of proteins with which they immunospecifically interact. Antibodies may also be used to immunoprecipitate proteins from a sample containing a mixture of proteins and other biological molecules. Other uses of anti-MED1 antibodies are described below.

Antibodies according to the present invention may be modified in a number of ways. Indeed the term "antibody" should be construed as covering any binding substance having a binding domain with the required

specificity. Thus, the invention covers antibody fragments, derivatives, functional equivalents and homologues of antibodies, including synthetic molecules and molecules whose shape mimics that of an antibody enabling it to bind an antigen or epitope.

Exemplary antibody fragments, capable of binding an antigen or other binding partner, are Fab fragment consisting of the VL, VH, Cl and CH1 domains; the Fd fragment consisting of the VH and CH1 domains; the Fv fragment consisting of the VL and VH domains of a single arm of an antibody; the dAb fragment which consists of a VH domain; isolated CDR regions and F(ab')₂ fragments, a bivalent fragment including two Fab fragments linked by a disulphide bridge at the hinge region. Single chain Fv fragments are also included.

Humanized antibodies in which CDRs from a non-human source are grafted onto human framework regions, typically with alteration of some of the framework amino acid residues, to provide antibodies which are less immunogenic than the parent non-human antibodies, are also included within the present invention.

II. Uses of MED1-Encoding Nucleic Acids, MED1 Proteins and Antibodies Thereto

MED1 appears to be an important DNA repair endonuclease which may play a role in mismatch repair. Mutations in MED1 are associated with certain forms of colon and endometrial cancer. The MED1 molecules of the invention may be used to advantage in genetic screening assays to identify those patients that may be at risk. Screening assays may also be developed which assess aberrant MED1 activity associated with Fragile X syndrome and other diseases characterized by triplet repeat expansion. Due to its methyl-CpG binding domain, MED1 might be useful in the analysis of genome methylation and of methylation-mediated DNA transcription, replication and repair (for instance, by

cleaving methylated and non-methylated DNA in a differential manner). Due to its endonuclease activity, MED1 is expected to be useful in the context of DNA manipulation technology. The employment of MED1 would
5 be of particular interest in the area of mutation detection. Other endonucleases have been successfully used to detect mutations based on recognition of cleavage products of heteroduplex intermediates carrying mismatches (Mashal R.D., Koontz J. and Sklaar J. Nature
10 Genet. 9: 177-183, 1995; Smith J. and Modrich P. Proc. Natl. Acad. Sci USA 93: 4374-4379, 1996).

Additionally, MED1 nucleic acids, proteins and antibodies thereto, according to this invention, may be used as a research tool to identify other proteins that
15 are intimately involved in DNA recognition and repair reactions. Biochemical elucidation of the DNA recognition and repair capacity of MED1 will facilitate the development of these novel screening assays for assessing a patient's propensity for cancer and genetic
20 disease.

A. MED1-Encoding Nucleic Acids

MED1-encoding nucleic acids may be used for a variety of purposes in accordance with the present
25 invention. MED1-encoding DNA, RNA, or fragments thereof may be used as probes to detect the presence of and/or expression of genes encoding MED1 proteins. Methods in which MED1-encoding nucleic acids may be utilized as probes for such assays include, but are not limited to:
30 (1) *in situ* hybridization; (2) Southern hybridization (3) northern hybridization; and (4) assorted amplification reactions such as polymerase chain reactions (PCR).

The MED1-encoding nucleic acids of the
35 invention may also be utilized as probes to identify related genes from other animal species. As is well known in the art, hybridization stringencies may be

adjusted to allow hybridization of nucleic acid probes with complementary sequences of varying degrees of homology. Thus, MED1-encoding nucleic acids may be used to advantage to identify and characterize other genes of varying degrees of relation to MED1, thereby enabling further characterization of the DNA repair system. Additionally, they may be used to identify genes encoding proteins that interact with MED1 (e.g., by the "interaction trap" technique), which should further accelerate identification of the components involved in DNA repair.

Nucleic acid molecules, or fragments thereof, encoding MED1 may also be utilized to control the production of MED1, thereby regulating the amount of protein available to participate in DNA repair reactions. Alterations in the physiological amount of MED1 protein may dramatically affect the activity of other protein factors involved in DNA repair.

The availability of MED1 encoding nucleic acids enables the production of strains of laboratory mice carrying part or all of the MED1 gene or mutated sequences thereof. Such mice may provide an in vivo model for cancer. Alternatively, the MED1 sequence information provided herein enables the production of knockout mice in which the endogenous gene encoding MED1 has been specifically inactivated. Methods of introducing transgenes in laboratory mice are known to those of skill in the art. Three common methods include: 1. integration of retroviral vectors encoding the foreign gene of interest into an early embryo; 2. injection of DNA into the pronucleus of a newly fertilized egg; and 3. the incorporation of genetically manipulated embryonic stem cells into an early embryo. Production of the transgenic mice described above will facilitate the molecular elucidation of the role MED1 plays in embryonic development and cancer.

A transgenic mouse carrying the human MED1 gene is

generated by direct replacement of the mouse MED1 gene with the human gene. These transgenic animals are useful for drug screening studies as animal models for human diseases and for eventual treatment of disorders or diseases associated with biological activities modulated by MED1. A transgenic animal carrying a "knock out" of MED1 is useful for assessing the role of MED1 in maintaining DNA fidelity.

As a means to define the role that MED1 plays in mammalian systems, mice may be generated that cannot make MED1 protein because of a targeted mutational disruption of the MED1 gene.

The term "animal" is used herein to include all vertebrate animals, except humans. It also includes an individual animal in all stages of development, including embryonic and fetal stages. A "transgenic animal" is any animal containing one or more cells bearing genetic information altered or received, directly or indirectly, by deliberate genetic manipulation at the subcellular level, such as by targeted recombination or microinjection or infection with recombinant virus. The term "transgenic animal" is not meant to encompass classical cross-breeding or in vitro fertilization, but rather is meant to encompass animals in which one or more cells are altered by or receive a recombinant DNA molecule. This molecule may be specifically targeted to a defined genetic locus, be randomly integrated within a chromosome, or it may be extrachromosomally replicating DNA. The term "germ cell line transgenic animal" refers to a transgenic animal in which the genetic alteration or genetic information was introduced into a germ line cell, thereby conferring the ability to transfer the genetic information to offspring. If such offspring, in fact, possess some or all of that alteration or genetic information, then they, too, are transgenic animals.

The alteration or genetic information may be

foreign to the species of animal to which the recipient belongs, or foreign only to the particular individual recipient, or may be genetic information already possessed by the recipient. In the last case, the
5 altered or introduced gene may be expressed differently than the native gene.

The altered MED1 gene generally should not fully encode the same MED1 protein native to the host animal and its expression product should be altered to a minor
10 or great degree, or absent altogether. However, it is conceivable that a more modestly modified MED1 gene will fall within the compass of the present invention if it is a specific alteration.

The DNA used for altering a target gene may be
15 obtained by a wide variety of techniques that include, but are not limited to, isolation from genomic sources, preparation of cDNAs from isolated mRNA templates, direct synthesis, or a combination thereof.

A type of target cell for transgene introduction is
20 the embryonal stem cell (ES). ES cells may be obtained from pre-implantation embryos cultured in vitro (Evans et al., (1981) Nature 292:154-156; Bradley et al., (1984) Nature 309:255-258; Gossler et al., (1986) Proc. Natl. Acad. Sci. 83:9065-9069). Transgenes can be
25 efficiently introduced into the ES cells by standard techniques such as DNA transfection or by retrovirus-mediated transduction. The resultant transformed ES cells can thereafter be combined with blastocysts from a non-human animal. The introduced ES cells thereafter
30 colonize the embryo and contribute to the germ line of the resulting chimeric animal.

One approach to the problem of determining the contributions of individual genes and their expression products is to use isolated MED1 genes to selectively
35 inactivate the wild-type gene in totipotent ES cells (such as those described above) and then generate transgenic mice. The use of gene-targeted ES cells in

the generation of gene-targeted transgenic mice was described, and is reviewed elsewhere (Frohman et al., (1989) Cell 56:145-147; Bradley et al., (1992) Bio/Technology 10:534-539).

5 Techniques are available to inactivate or alter any genetic region to a mutation desired by using targeted homologous recombination to insert specific changes into chromosomal alleles. However, in comparison with homologous extrachromosomal recombination, which occurs
10 at a frequency approaching 100%, homologous plasmid-chromosome recombination was originally reported to only be detected at frequencies between 10^{-6} and 10^{-3} . Nonhomologous plasmid-chromosome interactions are more frequent occurring at levels 10^5 -fold to 10^2 -fold greater
15 than comparable homologous insertion.

 To overcome this low proportion of targeted recombination in murine ES cells, various strategies have been developed to detect or select rare homologous recombinants. One approach for detecting homologous
20 alteration events uses the polymerase chain reaction (PCR) to screen pools of transformant cells for homologous insertion, followed by screening of individual clones. Alternatively, a positive genetic selection approach has been developed in which a marker
25 gene is constructed which will only be active if homologous insertion occurs, allowing these recombinants to be selected directly. One of the most powerful approaches developed for selecting homologous recombinants is the positive-negative selection (PNS)
30 method developed for genes for which no direct selection of the alteration exists. The PNS method is more efficient for targeting genes which are not expressed at high levels because the marker gene has its own promoter. Non-homologous recombinants are selected
35 against by using the Herpes Simplex virus thymidine kinase (HSV-TK) gene and selecting against its nonhomologous insertion with effective herpes drugs such

as gancyclovir (GANC) or (1-(2-deoxy-2-fluoro-B-D arabinofluranosyl)-5-iodouracil, (FIAU). By this counter selection, the number of homologous recombinants in the surviving transformants can be increased.

5 As used herein, a "targeted gene" or "knock-out" is a DNA sequence introduced into the germline or a non-human animal by way of human intervention, including but not limited to, the methods described herein. The targeted genes of the invention include DNA sequences
10 which are designed to specifically alter cognate endogenous alleles.

Methods of use for the transgenic mice of the invention are also provided herein. Therapeutic agents for the treatment or prevention of cancer may be
15 screened in studies using MED1 transgenic mice.

In another embodiment of the invention, MED1 knockout mice may be used to produce an array of monoclonal antibodies specific for MED1 protein.

As described above, MED1-encoding nucleic acids are
20 also used to advantage to produce large quantities of substantially pure MED1 protein, or selected portions thereof.

B. MED1 Protein and Antibodies

25 Purified MED1, or fragments thereof, may be used to produce polyclonal or monoclonal antibodies which also may serve as sensitive detection reagents for the presence and accumulation of MED1 (or complexes containing MED1) in mammalian cells. Recombinant
30 techniques enable expression of fusion proteins containing part or all of the MED1 protein. The full length protein or fragments of the protein may be used to advantage to generate an array of monoclonal antibodies specific for various epitopes of the protein,
35 thereby providing even greater sensitivity for detection of the protein in cells.

Polyclonal or monoclonal antibodies

immunologically specific for MED1 may be used in a variety of assays designed to detect and quantitate the protein. Such assays include, but are not limited to: (1) flow cytometric analysis; (2) immunochemical localization of MED1 in tumor cells; and (3) immunoblot analysis (e.g., dot blot, Western blot) of extracts from various cells. Additionally, as described above, anti-MED1 can be used for purification of MED1 (e.g., affinity column purification, immunoprecipitation).

From the foregoing discussion, it can be seen that MED1-encoding nucleic acids, MED1 expressing vectors, MED1 proteins and anti-MED1 antibodies of the invention can be used to detect MED1 gene expression and alter MED1 protein accumulation for purposes of assessing the genetic and protein interactions involved in the recognition and repair of DNA damage.

Exemplary approaches for detecting MED1 nucleic acid or polypeptides/proteins include:

a) comparing the sequence of nucleic acid in the sample with the MED1 nucleic acid sequence to determine whether the sample from the patient contains mutations; or

b) determining the presence, in a sample from a patient, of the polypeptide encoded by the MED1 gene and, if present, determining whether the polypeptide is full length, and/or is mutated, and/or is expressed at the normal level; or

c) using DNA restriction mapping to compare the restriction pattern produced when a restriction enzyme cuts a sample of nucleic acid from the patient with the restriction pattern obtained from normal MED1 gene or from known mutations thereof; or,

d) using a specific binding member capable of binding to a MED1 nucleic acid sequence (either normal sequence or known mutated sequence), the specific binding member comprising nucleic acid hybridizable with the MED1 sequence, or substances comprising an antibody

domain with specificity for a native or mutated MED1 nucleic acid sequence or the polypeptide encoded by it, the specific binding member being labelled so that binding of the specific binding member to its binding partner is detectable; or,

5 e) using PCR involving one or more primers based on normal or mutated MED1 gene sequence to screen for normal or mutant MED1 gene in a sample from a patient.

A "specific binding pair" comprises a specific binding member (sbm) and a binding partner (bp) which have a particular specificity for each other and which in normal conditions bind to each other in preference to other molecules. Examples of specific binding pairs are antigens and antibodies, ligands and receptors and complementary nucleotide sequences. The skilled person is aware of many other examples and they do not need to be listed here. Further, the term "specific binding pair" is also applicable where either or both of the specific binding member and the binding partner comprise a part of a large molecule. In embodiments in which the specific binding pair are nucleic acid sequences, they will be of a length to hybridize to each other under conditions of the assay, preferably greater than 10 nucleotides long, more preferably greater than 15 or 20 nucleotides long.

In most embodiments for screening for cancer susceptibility alleles, the MED1 nucleic acid in the sample will initially be amplified, e.g. using PCR, to increase the amount of the analyte as compared to other sequences present in the sample. This allows the target sequences to be detected with a high degree of sensitivity if they are present in the sample. This initial step may be avoided by using highly sensitive array techniques that are becoming increasingly important in the art.

The identification of the MED1 gene and its association with cancer paves the way for aspects of the

present invention to provide the use of materials and methods, such as are disclosed and discussed above, for establishing the presence or absence in a test sample of a variant form of the gene, in particular an allele or variant specifically associated with cancer, especially colorectal or endometrial cancer. This may be for diagnosing a predisposition of an individual to cancer. It may be for diagnosing cancer of a patient with the disease as being associated with the gene.

This allows for planning of appropriate therapeutic and/or prophylactic measures, permitting stream-lining of treatment. The approach further stream-lines treatment by targeting those patients most likely to benefit.

According to another aspect of the invention, methods of screening drugs for cancer therapy to identify suitable drugs for restoring MED1 product functions are provided. A major problem in cancer treatment is the development of drug resistance or ionizing radiation resistance by the tumor cells which eventually leads to failure of therapy. Recent studies have revealed that inactivation of DNA mismatch repair is an important mechanism of resistance to many chemotherapeutic drugs used in the clinic (Fink D., Aebi S. and Howell S.B. (1998). Clinical Cancer Res. 4: 1-6). In fact, a functional mismatch repair system appears to be required for killing by many alkylating agents and platinum compounds. Resistance/tolerance to those agents is associated with loss of expression or function of mismatch repair genes: in the absence of a functional mismatch repair system, DNA damage accumulates but fails to trigger apoptosis (Fink D., Aebi S. and Howell S.B. (1998), *supra*). Defects in DNA mismatch repair genes (hMLH1, hPMS2, hMSH2 and hMSH6) have been found in cell lines and primary tumors resistant to those chemotherapeutic agents. Thus, loss of MED1 function/expression may be associated with tumor drug

resistance. Restoring of MED1 function by gene transfer or by pharmacological means would be expected to overcome resistance to treatment.

5 The MED1 polypeptide or fragment employed in drug screening assays may either be free in solution, affixed to a solid support or within a cell. One method of drug screening utilizes eukaryotic or prokaryotic host cells which are stably transformed with recombinant polynucleotides expressing the polypeptide or fragment, 10 preferably in competitive binding assays. Such cells, either in viable or fixed form, can be used for standard binding assays. One may determine, for example, formation of complexes between a MED1 polypeptide or fragment and the agent being tested, or examine the 15 degree to which the formation of a complex between a MED1 polypeptide or fragment and a known ligand is interfered with by the agent being tested.

Another technique for drug screening provides high throughput screening for compounds having suitable 20 binding affinity to the MED1 polypeptides and is described in detail in Geysen, PCT published application WO 84/03564, published on Sep. 13, 1984. Briefly stated, large numbers of different, small peptide test compounds are synthesized on a solid substrate, such as 25 plastic pins or some other surface. The peptide test compounds are reacted with MED1 polypeptide and washed. Bound MED1 polypeptide is then detected by methods well known in the art.

Purified MED1 can be coated directly onto plates 30 for use in the aforementioned drug screening techniques. However, non-neutralizing antibodies to the polypeptide can be used to capture antibodies to immobilize the MED1 polypeptide on the solid phase.

35 This invention also contemplates the use of competitive drug screening assays in which neutralizing antibodies capable of specifically binding the MED1 polypeptide compete with a test compound for binding to

the MED1 polypeptide or fragments thereof. In this manner, the antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants of the MED1 polypeptide.

5 A further technique for drug screening involves the use of host eukaryotic cell lines or cells (such as described above) which have a nonfunctional MED1 gene. These host cell lines or cells are defective at the MED1 polypeptide level. The host cell lines or cells are
10 grown in the presence of drug compound. The rate of growth of the host cells is measured to determine if the compound is capable of regulating the growth of MED1 defective cells.

 The goal of rational drug design is to produce
15 structural analogs of biologically active polypeptides of interest or of small molecules with which they interact (e.g., agonists, antagonists, inhibitors) in order to fashion drugs which are, for example, more active or stable forms of the polypeptide, or which,
20 e.g., enhance or interfere with the function of a polypeptide *in vivo*. See, e.g., Hodgson, (1991) Bio/Technology 9:19-21. In one approach, one first determines the three-dimensional structure of a protein of interest (e.g., MED1 polypeptide) or, for example, of
25 the MED1-DNA complex, by x-ray crystallography, by nuclear magnetic resonance, by computer modeling or most typically, by a combination of approaches. Less often, useful information regarding the structure of a polypeptide may be gained by modeling based on the
30 structure of homologous proteins. An example of rational drug design is the development of HIV protease inhibitors (Erickson et al., (1990) Science 249:527-533). In addition, peptides (e.g., MED1 polypeptide) may be analyzed by an alanine scan (Wells, 1991) Meth.
35 Enzym. 202:390-411. In this technique, an amino acid residue is replaced by Ala, and its effect on the peptide's activity is determined. Each of the amino acid

residues of the peptide is analyzed in this manner to determine the important regions of the peptide.

It is also possible to isolate a target-specific antibody, selected by a functional assay, and then to solve its crystal structure. In principle, this approach yields a pharmacore upon which subsequent drug design can be based. It is possible to bypass protein crystallography altogether by generating anti-idiotypic antibodies (anti-ids) to a functional, pharmacologically active antibody. As a mirror image of a mirror image, the binding site of the anti-ids would be expected to be an analog of the original molecule. The anti-id could then be used to identify and isolate peptides from banks of chemically or biologically produced banks of peptides. Selected peptides would then act as the pharmacore.

Thus, one may design drugs which have, e.g., improved MED1 polypeptide activity or stability or which act as inhibitors, agonists, antagonists, etc. of MED1 polypeptide activity. By virtue of the availability of cloned MED1 sequences, sufficient amounts of the MED1 polypeptide may be made available to perform such analytical studies as x-ray crystallography. In addition, the knowledge of the MED1 protein sequence provided herein will guide those employing computer modeling techniques in place of, or in addition to x-ray crystallography.

III Therapeutics

A. Pharmaceuticals and Peptide Therapies

The MED1 polypeptides/proteins, antibodies, peptides and nucleic acids of the invention can be formulated in pharmaceutical compositions. These compositions may comprise, in addition to one of the above substances, a pharmaceutically acceptable excipient, carrier, buffer, stabilizer or other materials well known to those skilled in the art. Such

materials should be non-toxic and should not interfere with the efficacy of the active ingredient. The precise nature of the carrier or other material may depend on the route of administration, e.g. oral, intravenous, cutaneous or subcutaneous, nasal, intramuscular, intraperitoneal routes.

Whether it is a polypeptide, antibody, peptide, nucleic acid molecule, small molecule or other pharmaceutically useful compound according to the present invention that is to be given to an individual, administration is preferably in a "prophylactically effective amount" or a "therapeutically effective amount" (as the case may be, although prophylaxis may be considered therapy), this being sufficient to show benefit to the individual.

B. Methods of Gene Therapy

As a further alternative, the nucleic acid encoding the authentic biologically active MED1 polypeptide could be used in a method of gene therapy, to treat a patient who is unable to synthesize the active "normal" polypeptide or unable to synthesize it at the normal level, thereby providing the effect elicited by wild-type MED1 and suppressing the occurrence of "abnormal" MED1 lacking the ability to perform or effect DNA repair.

Vectors such as viral vectors have been used in the prior art to introduce genes into a wide variety of different target cells. Typically the vectors are exposed to the target cells so that transformation can take place in a sufficient proportion of the cells to provide a useful therapeutic or prophylactic effect from the expression of the desired polypeptide. The transfected nucleic acid may be permanently incorporated into the genome of each of the targeted tumor cells, providing long lasting effect, or alternatively the treatment may have to be repeated periodically.

A variety of vectors, both viral vectors and plasmid vectors are known in the art, see US Patent No. 5,252,479 and WO 93/07282. In particular, a number of viruses have been used as gene transfer vectors, including papovaviruses, such as SV40, vaccinia virus, herpes viruses including HSV and EBV, and retroviruses. Many gene therapy protocols in the prior art have employed disabled murine retroviruses.

Gene transfer techniques which selectively target the MED1 nucleic acid to colorectal tissues are preferred. Examples of this include receptor-mediated gene transfer, in which the nucleic acid is linked to a protein ligand via polylysine, with the ligand being specific for a receptor present on the surface of the target cells.

The following examples are provided to illustrate certain embodiments of the invention. They are not intended to limit the invention in any way.

EXAMPLE I

The methods described below have been used to advantage to isolate the MED1 encoding nucleic acids of the invention.

A. Interaction trap screen, cDNA and genomic DNA isolation.

Yeast interaction trap screening (Gyuris et al., (1993) *Cell* 75:791-803; Golemis et al., (1996) Yeast Interaction Trap/Two Hybrid Systems to Identify Interacting Proteins, Unit 20.1.1-20.1.28 in *Current Protocols in Molecular Biology*, eds. Ausubel, F.M. et al., John Wiley & Sons, NY) was used to isolate cDNAs encoding proteins able to interact with hMLH1. The hMLH1 open reading frame was inserted into the polylinker of the pEG202 vector (Golemis et al., 1996, supra). The resulting "bait" construct pEG202-t-hMLH1 expresses the hMLH1 protein (amino acids 1-756) as a carboxyterminal fusion to the LexA DNA binding protein. *Saccharomyces cerevisiae* strain EGY191 (Estojak et al.,

(1995) *Mol. Cell Bio.* 15:5820-5829) was transformed with the bait construct and with the LacZ reporter plasmid pSH18-34 (Golemis et al., 1996, supra).

5 The EGY191/pSH18-34/pEG202-t-hMLH1 cells were supertransformed with a human fetal brain cDNA library constructed in the vector pJG4-5. This vector directs the synthesis of proteins fused to the B42 transcriptional activator domain (Ruden et al., (1991) *Nature* 350:25-252) and the expression is controlled by
10 the galactose-inducible GAL1 promoter. Approximately 4×10^5 independent transformants were obtained in yeast and used for screening. For selection of the positive interactors, the supertransformed cells were cultured on leucine-minus / galactose solid medium. Colonies
15 growing on this medium after 3-5 days incubation were subcultured on leucine-minus or X-Gal media containing either glucose or galactose as a carbon source. Twenty-two colonies growing on leucine-minus / galactose but not leucine-minus / glucose medium and turning blue
20 on X-Gal / galactose but not X-Gal / glucose plates were further characterized.

Plasmid DNA encoding putative hMLH1 interactors was isolated from these colonies (clones f1 through f22), transferred first to KC8 and then to XL-1 blue *E. coli*
25 strains, and sequenced. These and subsequent sequencing reactions were performed on double stranded DNA with the ABI automated sequencer 377 using dye terminator chemistry (Perkin Elmer). Sequence assembling and analysis was performed with the Genetics Computer Group
30 software (Genetics Computer Group, 1994). Since the f5 clone (later named MED1) was shorter (0.8 kb 3' of B42) than the mRNA transcript detected in human tissues by Northern blot analysis (approximately 2.4 kb), a f5-derived probe was used to screen three additional
35 cDNA lambda libraries. The libraries, derived from human fetal brain (Stratagene and Clontech) and from the ovarian cancer cell line C200 (gift of Drs. A. Godwin

and G. Kruh), were screened following standard procedures as previously described (Bellacosa et al., 1994, supra).

Screening of a human genomic DNA library prepared in the lambda phage FIX II (Stratagene) with the f5/MED1 cDNA probe yielded six clones. One of these clones (# 16) was further characterized and subcloned in plasmid vectors. Sequence analysis of the subclones and comparison to the MED1 cDNA sequence allowed mapping of seven MED1 exons (exons 2 through 8, Fig. 14). The remaining exon (exon 1) and the intervening intron between exon 1 and exon 2 was cloned by PCR utilizing human genomic DNA as template and the primers of Sequence I.D. No. 6 and 20. SEQ ID NO: 20 is CAAATCTTCCTGCTGTCTTCC which maps within exon 2. Table I provides suitable primer sets for amplifying exons of the MED1 gene.

This human genomic clone has been deposited with the American Type Culture Collection, 10801 University Blvd., Manassas, VA 20110-2209 on July 28, 1998 under the terms of the Budapest Treaty, Accession Number: Not yet assigned. The sequence of the human genomic clone is shown in Figure 20, SEQ ID NO:22.

TABLE I. OLIGONUCLEOTIDE PRIMERS FOR MED1

	5' primer	3' primer
exon 1	GTCTGGGGCGCTTTCCGAA (SEQ ID NO:6)	CCACACACTGTCCACTCTCCCG (SEQ ID NO:7)
exon 2	ACTCCCATAGCACAAGACTGG (SEQ ID NO:8)	GCTATGCTCCCACTACCTGC (SEQ ID NO:9)
exon 3	CCCTTCTATTTACTAGCAGTA (SEQ ID NO:10)	GATGCAGCATATAAATTTCTC (SEQ ID NO:11)
exons 4 and 5	TGCATCCCTCAATATTGCTTT (SEQ ID NO:12)	TCAATTCAGTGCTTTCTCCCT (SEQ ID NO:13)
exon 6	AGCCACCTGGAGTCTTGTA (SEQ ID NO:14)	AAAGTTTAAGGTGTGGCTCTC (SEQ ID NO:15)
exon 7	GAAGCTGACCTGATAATGTGG (SEQ ID NO:16)	CTTATTTGCCTCAGAGACCA (SEQ ID NO:17)
exon 8	TATCGTAATGTACTGTCCCCC (SEQ ID NO:18)	GCTTAGCAAGGCTGATAGAA (SEQ ID NO:19)

Screening at low stringency of a mouse 129/SVJ strain genomic DNA library prepared in the lambda phage FIX II (Stratagene) with the same HindIII-HindIII

fragment derived from the human MED1 cDNA probe (from nucleotide 1513-1935 of SEQ ID NO: 1) yielded one clone. This clone (#3) was further characterized and subcloned in plasmid vectors. Sequence analysis of the subclones and comparison to the human MED1 cDNA and genomic sequence allowed mapping of seven mouse MED1 exons (exons 1 through 7. Fig. 16). Assembling of the mouse MED1 exons allowed the derivation of a partial sequence of the mouse MED1 cDNA (Fig. 17). From the latter sequence a partial predicted amino acid sequence of the mouse MED1 protein was derived and it was shown to be highly conserved by comparison to the human MED1 protein sequence (Fig. 18). This mouse genomic clone has been deposited with the American Type Culture Collection, 10801 University Blvd., Manassas, VA 20110-2209 on July 28, 1998 under the terms of the Budapest Treaty, Accession Number: Not yet assigned. The sequence of the mouse genomic clone is shown in Figure 19, SEQ ID NO:21.

B. Northern and Southern blot analysis.

A multiple tissue northern blot of poly-A selected RNA (Clontech) was hybridized under high-stringency conditions to a ³²P-labeled 0.8 kb f5 probe. The blot was washed to a final stringency of 0.1 x SSC/0.1% SDS (1 x SSC is 0.15 M NaCl/0.015 M sodium citrate) at 65°C for 40 minutes, and then exposed to X-ray film (Kodak X-Omat AR) at -70°C.

For the "Zoo" blot experiment, genomic DNA prepared from vertebrate species was digested with the restriction enzyme HindIII (New England Biolabs), separated on a 0.8% agarose gel and transferred to a nylon membrane. The membrane was hybridized to a ³²P-labelled human MED1 cDNA probe (HindIII-HindIII fragment from nucleotide 1513 to nucleotide 1935 of the Sequence I.D. No. 1). Hybridization was performed in a solution containing 35% formamide, 6x SSC, 5x Denhardt's solution, 20 mM sodium phosphate pH 6.5, 20 micrograms/ml of sheared E. coli genomic DNA and 0.5% sodium dodecyl

sulfate (SDS). The filter was washed twice at room temperature and twice at 65°C in a solution containing 4x SSC and 0.1% SDS. Hybridization signals were revealed by autoradiography.

- 5 Hybridization of the HindIII-HindIII fragment probe (from nucleotide 1513 to nucleotide 1935 of the Sequence I.D. No. 1) at low stringency to a "zoo" blot revealed conservation of the MED1 gene among vertebrates. See Figure 15.

10 **C. In vitro transcription and translation.**

Coupled in vitro transcription and translation was conducted with a rabbit reticulocyte lysate- and T7 RNA polymerase-based kit (Promega), following the manufacturer's recommendations and employing

15 ³⁵S-methionine (Amersham).

D. Cell culture, expression constructs, and transfections.

NIH 3T3 cells were cultured in Dulbecco's modified Eagle's medium supplemented with 10% calf serum, penicillin (50 units/ml), streptomycin (50 µg/ml), and kanamycin (100 µg/ml). The expression constructs of MED1 were generated in the CMV promoter-based CMV5 vector, a derivative of CMV4 (Andersson et al., (1989) *J. Biol. Chem.* 264:8222-8229). For construction of the hemagglutinin epitope carboxy-terminally tagged MED1

20 plasmid, the MED1 cDNA was inserted in place of the Gfi-1 ZN mutant construct open reading frame (Grimes et al. (1996) *Mol. Cell Bio.* 16:6263-6272), a gift of Dr. Leighton Grimes. For construction of the hemagglutinin

25 epitope amino terminally-tagged MED1 plasmids M1 and M2, a XbaI site was inserted by polymerase chain reaction immediately upstream of the ATG codons at nucleotide positions 142 and 262, respectively. Then the MED1 open

30 reading frame, excised with XbaI and NsiI (blunted), was inserted in place of the Akt gene in the CMV5

35 hemagglutinin tag·Akt construct (Datta et al., (1996) *J. Biol. Chem.* 271:30835-30839).

Transient transfections of NIH 3T3 cells seeded in 6-well plates at 0.15×10^6 cells/well were carried out using 1.5 μ g of DNA and 6 μ l of lipofectamine (Life Technologies, Inc.), following the manufacturer's protocol. Forty-eight hours after transfection, cells were washed twice with Dulbecco phosphate buffered saline and then lysed with RIPA buffer (10mM sodium phosphate pH 7.0, 150mM NaCl, 1% w/v sodium deoxycholate, 1% v/v Nonidet P-40, 0.1% w/v sodium dodecylsulfate, 1mM phenylmethanesulfonyl-fluoride, 2 μ g/ml aprotinin, 2 μ g/ml leupeptin, 50mM NaF, 1mM sodium pyrophosphate, 1mM sodium orthovanadate, 1mM dithiothreitol, and 2mM EDTA).

E. Western blotting.

Cell lysates were separated by sodium dodecylsulfate-polyacrylamide gel electrophoresis (SDS-PAGE) in 8.5% gels and transferred to Immobilon P membranes (Millipore) by electroblotting with a Genie apparatus (Idea Scientific Co.) in a buffer containing 25mM Tris-HCl pH 8, 190mM glycine and 20% v/v methanol. Following overnight incubation in 5% dry milk in Tris-buffered saline (TBS: 0.9% w/v NaCl, 10mM Tris-HCl pH 7.4, 0.05% w/v $MgCl_2$), the membrane was incubated for 1 hour at room temperature with the anti-hemagglutinin tag monoclonal antibody 12CA5 (Boehringer) in 2% dry milk in TBS. After three 10-minute washes in TBS supplemented with 0.1% v/v Tween-20, the membrane was incubated for 40 minutes at room temperature with an anti-mouse secondary antibody conjugated to horseradish peroxidase (Amersham). Following washing, the bound secondary antibody was detected by enhanced chemiluminescence (Amersham).

F. Fluorescence in situ hybridization.

Metaphase spreads from normal human lymphocytes were prepared according to published methods (Fan et al. (1990) *Proc. Natl. Acad. Sci.* 87:6223-6227). Nick translation was used to label a MED1 genomic DNA

subclone with biotin-16-dUTP. Three hundred ng of the probe were then mixed with 150 μ g of human Cot-1 DNA (Life Technologies Inc.) and 50 μ g salmon sperm DNA to block repetitive elements. The DNA was denatured at 75°C for 5 minutes and then reannealed for 1 hour at 37°C prior to hybridization to metaphase spreads overnight at 37°C. The MED1 signal was detected with fluorescein isothiocyanate-labeled avidin (Oncor), whereas the chromosomes were counterstained with propidium iodide (Oncor). Metaphase spreads were observed using a Zeiss Axiophot microscope and images were captured by a cooled CCD camera (Photometrics) connected to a computer workstation. To identify the precise chromosomal location of the probe, the separate digitized images of FITC and propidium iodide were merged using Oncor version 1.6 software.

G. Electromobility shift analysis

Transient transfections of 293 cells seeded in 10-cm dishes were carried out using 12 μ g of DNA and 48 μ l lipofectamine (Life Technologies, Inc.), following the manufacturer's protocol. Seventy-two hours after transfections, cells were washed twice with Dulbecco's phosphate buffered saline and then lysed with NP-40 lysis buffer (0.5% Nonidet P-40, 10% glycerol, 137 mM NaCl, 20 mM Tris-HCl, pH 7.4) containing 1 mM phenylmethylsulfonylfluoride, 2 μ g/ml aprotinin, 2 μ g/ml leupeptin, 1 mM NaF, 1 mM sodium pyrophosphate, 1 mM sodium orthovanadate, and 1 mM dithiothreitol. Nuclei were disrupted by sonication with a sonic dismembrator (Fisher). Flag-MED1 was immunoprecipitated from the cell lysates with an anti-Flag antibody coupled to agarose beads (Kodak) and then eluted in a 50 μ l volume with a solution containing a molar excess of Flag-peptide (Kodak) in electromobility shift analysis (EMSA) buffer (10 mM Tris-HCl, pH 7.5, 50 mM NaCl, 0.5 mM EDTA, 5% glycerol). A double stranded oligonucleotide containing five fully methylated CpG sites was generated

by annealing the following oligonucleotides (M= 5-methylcytosine):

Sequence I. D. No. 3:

5'-GCCAATTCMGTGCGAMGAAGCMGGACGATMGACCAGMGCTCGAGCA-3'

5 Sequence I. D. No. 4:

5'-GTGCTCGAGMGCTGGTMGATCGTCMGGCTTMGTGCGCAMGGAATTCG-3'

The double-stranded oligonucleotide was labeled with ³²P- α -dCTP and Klenow enzyme. EMSA was conducted as described previously (Durand et al., (1988) *Mol. Cell. Biol.* 8:1715-1724). Briefly, binding of MED1 to labeled oligonucleotides was carried out by incubating 1 μ l out of 50 μ l of the MED1 eluate, 7 X 10⁴ cpm of labeled oligonucleotides and 4 μ g of poly (dI-dC) in EMSA buffer (final volume of 20 μ l) at room temperature. Competition was carried out in the presence of 100 ng (100-fold excess) of the cold oligonucleotide. Binding reactions were separated on a 6% non-denaturing polyacrylamide gel and visualized by autoradiography of the dried gel.

20 For the electromobility shift assay employing the purified methyl-CpG binding domain (MBD) of MED1, the methylated probe was assembled by annealing the two complementary oligonucleotides of Sequence I.D. No. 3 and Sequence I.D. No. 4. containing 5-methylcytosine. 25 See Figure 10B. The unmethylated probe was assembled with two complementary oligonucleotides of identical sequence to the oligonucleotides of Sequence I.D. No. 3 and Sequence I.D. No. 4., except that cytosine replaced 5-methylcytosine. Labeling of the probes was conducted 30 as above. DNA binding reactions were carried out in 10 mM Tris-HCl pH 7.5, 50 mM NaCl, 5% glycerol, 0.5 mM EDTA, 0.5 mM DTT, in the presence of 0.5 mg of polydA/polydT (ICN) as non-specific competitor DNA [S. Buratowski and L.A. Chodosh, In *Current Protocols in Molecular Biology*, eds. F. M. Ausubel, et al., John Wiley & Sons, New York (1996)]. Bacterially expressed and purified MBD (20 ng) was incubated with the

³²P-labeled double-strand oligonucleotides (20,000 cpm, 0.2 ng) on ice for 30 min. For competition, the MBD was pre-incubated on ice for 20 min with a 100-fold excess of the cold oligonucleotide (20 ng) prior to addition of the probe. Binding reactions were loaded on a 10 % acrylamide gel and run at 4°C in 0.5x TBE. Dried gels were exposed to autoradiography.

H. Co-immunoprecipitation analysis

To analyze the interaction of MED1 with hMSH2, following transient transfection of 293 cells with the constructs of the invention, and lysis of cells after a 72 hour period, proteins were immunoprecipitated with anti-Flag antibody as described above, using an antibody against hMSH2. Immunoprecipitates were resuspended in Laemmli buffer, boiled for 10 minutes, separated on 8.5% SDS-PAGE and transferred to Immobilon P membranes. Western blotting was carried out as described above.

For analysis of the interaction of MED1 with hMLH1, HEK-293 cells were cultured at 37°C and 7.5% CO₂ in Dulbecco's modified Eagle's minimum essential medium (DMEM) supplemented with 10% fetal calf serum, penicillin (50 units/ml), streptomycin (50 µg/ml), and kanamycin (100 µg/ml). Cells seeded in 100-mm Petri dishes were transfected using LipofectAMINE (Life Technologies, Inc.) according to the manufacturer's protocol. Seventy-two hours later, cells were lysed on ice in one of three lysis buffers, containing 0.5% Nonidet P-40 (NP-40) [K. Datta et al., *Mol. Cell. Biol.* 15: 2304-2310 (1995)], 0.2% NP-40 [W. Gu, K. Bhatia, I.T. Magrath, C.V. Dang, R. Dalla-Favera, *Science* 264: 251-254 (1994)], or 1% Triton X-100 [S. F. Law et al., *Mol. Cell. Biol.* 16: 3327-3337 (1996)]; NP-40 lysates were mildly sonicated using a sonic dismembrator (Fisher). Immunoprecipitations were carried out with the anti-hemagglutinin tag antibody HA.11 coupled to beads (Berkeley Antibody Company). Immune complexes were washed with lysis buffer, and the proteins were

resolved by 8.5% SDS-polyacrylamide gel electrophoresis (SDS-PAGE) and transferred to PVDF membranes (Immobilon P, Millipore) with an X-genie electroblotter (Idea Scientific). Membranes were probed with an anti-MLH1 antibody (Pharmingen) and the HA.11 antibody (Berkeley Antibody Company). Detection of antigen-bound antibody was carried out using enhanced chemiluminescence (ECL, Amersham), according to the manufacturer's protocol. See Figure 11C.

10 **I. Expression of the MED1 catalytic (endonuclease) domain in *E. coli***

The nucleic acid sequence encoding the catalytic domain of MED1 was cloned in the vector pET28b (Novagen) as a carboxyterminal fusion to a 6xHis tag for expression in *E. coli*. This construct was transferred to the *E. coli* strain BL21(DE3)pLysS. Overnight cultures were diluted 1:15 in fresh medium and incubated for one-hour in a 37°C incubator. Expression of the construct was induced by addition of 1 mM IPTG for an additional 3 hours at 37°C. Cells were then collected by centrifugation and lysed in Laemmli buffer. Lysates were boiled for 10 minutes and separated on 12% SDS-PAGE. Proteins were visualized by Coomassie blue staining.

25 **J. Activity staining of the MED1-endonuclease domain after sodium dodecyl sulfate-polyacrylamide gel electrophoresis**

Activity staining of MED1 was performed essentially as described by Blank et al. (Blank et al. (1982) Analytical Biochemistry 120: 267-275). Briefly, bacterial lysates expressing the MED1 catalytic domain were separated in SDS-polyacrylamide gels (12%) containing 0.15 mg/ml heat-denatured calf thymus DNA. Following electrophoresis, the gel was incubated in a buffer containing 10 mM Tris-HCl pH 7.4 and 25% isopropanol for one hour at room temperature with one change of buffer every twenty minutes. After the first

hour, the gel was immersed in a buffer containing 10 mM Tris-HCl pH 7.4 for an additional hour with buffer changes every twenty minutes. The gel was then immersed in a buffer containing 10 mM Tris-HCl, pH 7.4, 10 mM MgCl₂, 5 mM CaCl₂, 2 μ M ZnCl₂ for 16 hours at room temperature to allow digestion of DNA. DNA was visualized by staining the gel with 0.2% toluidine blue 0 in 10 mM Tris-HCl pH 7.4, followed by destaining in 10 mM Tris-HCl pH 7.4 for one hour at room temperature with one change of buffer every 20 minutes. Deoxyribonuclease activity results in a zone of clearing indicating reduced DNA staining (Blank et al., (1982) *supra*).

K. Endonuclease activity of recombinant wild-type MED1.

The entire wild-type MED1 (codons 1-580, wt) and a deletion mutant lacking the endonuclease domain (codons 1-454, Δ endo) were expressed in bacteria and purified by nickel-agarose chromatography. For bacterial expression, PCR-generated fragments corresponding to the entire MED1 open reading frame or to isolated domains were cloned in pET28(b) (Novagen) and propagated in *E. coli* strain XL-1 Blue (Stratagene). Constructs were sequenced with an automated DNA sequencer (ABI) to verify that unwanted mutations were not inadvertently introduced; and they were transferred into *E. coli* strain BL21(DE3)pLysS. These cells were grown to O.D.₆₀₀ = 0.4 and then induced with 1 mM IPTG at 37°C for 3 hours. Bacterial lysates were purified over a nickel-agarose column (Ni²⁺-NTA agarose, Qiagen). Increasing amounts of the wild-type and Δ endo mutant (22, 44, 87.5 and 175 ng) were incubated with 500 ng of the 3.9 kb supercoiled plasmid pCR2 (Invitrogen) at 37°C for 30 min in a buffer containing 20 mM Tris-HCl pH 7.5, 25 mM KCl and 10 mM MgCl₂. Reaction products were separated on a 1% agarose gel buffered in 1x TAE and containing 0.25 μ g/ml ethidium bromide.

Identification and Characterization of MED1

To facilitate efforts to identify eukaryotic

functional homologues of the *E. coli* Muth endonuclease, the yeast interaction trap assay, a cloning strategy which screens for protein-protein interactions in the yeast *S. cerevisiae* (Golemis et al., 1996, supra) was employed. This strategy was based on the rationale that the human mismatch repair endonuclease would interact with hMLH1, the human MutL homologue, in a comparable way to what is observed in bacteria where the endonuclease Muth interacts with MutL. The complete coding sequence of hMLH1 (amino acids 1-756) was fused to the carboxy terminus of the DNA binding domain of LexA. This construct ("bait") was introduced along with the appropriate reporter plasmid in the yeast strain EGY191. EGY191, which harbors only two LexA operators directing transcription of the chromosomal LEU2 gene, was used because in initial experiments, employing the standard EGY48 strain, the bait protein had constitutive transcriptional activity (data not shown). Western blot analysis with an anti-LexA antibody showed that pEG202-t-hMLH1 directs the synthesis of the expected size product for a LexA-hMLH1 bait protein in EGY191. In control experiments, performed following standard procedures, this protein was transported to the nucleus and did not activate transcription of the chromosomal LEU2 gene and of the episomal LacZ gene (data not shown). The EGY191/pSH18-34/pEG202-t-hMLH1 yeast cells were supertransformed with a human fetal brain cDNA library (approximately 4×10^5 recombinants) fused to the B42 portable activation domain, and colonies growing on selective leucine-minus plates in the presence of galactose but not glucose as carbon source were isolated. Twenty-two clones (f1 to f22) were selected encoding putative hMLH1 interactors. One clone, designated f5, (later named MED1) was identified which strongly interacted with hMLH1, based on the early appearance of colonies on selective leucine-minus/galactose plates and on the intensity of

color formation of colonies grown on indicator X-Gal/galactose plates. The specificity of the f5-hMLH1 interaction was assayed by supertransforming virgin EGY191/pSH18-34/pEG202-t-hMLH1 cells with f5 plasmid DNA. As a control, EGY191/pSH18-34 cells transformed with bait constructs of pEG202-bicoid, -MYC, -K-rev, and empty pEG202 vector, were also supertransformed with f5 DNA. Cells transformed with the combination of f5 and pEG202-t-hMLH1 grew on leucine-minus / galactose but not leucine-minus / glucose medium and turned blue on X-Gal / galactose but not X-Gal / glucose plates. Control cells failed to grow on leucine-minus / galactose and to turn blue on X-Gal / galactose plates, confirming specificity of the interaction between f5 and hMLH1 as shown in Figure 1.

Initial sequence analysis revealed that f5, which was represented only once in this group of 22 putative interactors, codes for a protein sharing homology with several bacterial endonucleases involved in DNA repair. Since the f5-encoded protein is a putative DNA repair enzyme, its expression is expected to be ubiquitous. A Northern blot containing mRNA from multiple tissues was probed with the entire 0.8 kb insert of the f5 clone. This analysis revealed that, consistent with a putative housekeeping role in DNA repair, the f5 gene is expressed in all normal tissues tested with a transcript of approximately 2.4 kb. See Figure 2.

In order to clone the remaining portion of the gene, a f5-derived probe was used to screen four additional cDNA libraries, three from fetal brain and one from the ovarian cancer cell line C200. Six clones were isolated from the fetal brain libraries and 11 from the C200 library. These clones were sequenced. Overlapping sequences were aligned until the nearly complete sequence of the gene was determined (2.1 kb). See Figure 3. The MED1 transcript contains an open reading frame of 1740 bases, preceded by an in-frame

stop codon, which predicts a protein of about 580 amino acids encoded by the sequence of Sequence I.D. No. 2. Slight sequence variations were observed between the cDNA clones analyzed. These are set forth below:

5

SEQUENCE VARIATIONS

1) Nucleotides 1325-1342: 18 nucleotides - GTGAGAAAATATTTCAAG - are either present (as in Sequence I.D. No. 1) or absent (as in Sequence I. D. No. 23) from the cDNA, therefore the 6 amino acids encoded by those
10 nucleotides (GEKIFQ) are either present (as in Sequence I. D. No. 2) or absent (as in Sequence I. D. No.24) in the predicted protein. This variation appears to originate from alternative usage of a splice donor site. In the genomic DNA sequence:

15

...GACTTCACTG**GT**GAGAAAATATTTCAAGGT...

If the second splice donor site (bold) is used, then the 18 nucleotides GTGAGAAAATATTTCAAG are incorporated in the mRNA; if the first splice donor site (underlined) is used, then the same 18 nucleotides are
20 spliced out and are not incorporated in the mRNA.

2) Nucleotide 1876: T (as in Sequence I.D. No. 1) or C (as in Sequence I. D. No. 25), therefore codon 579 is either TTA or CTA (no amino acid variation, since both
25 code for leucine).

3) Nucleotide 2042: C (as in Sequence I.D. No. 1) or T (as in Sequence I. No. 26), (no amino acid variation, since this change is in the 3' untranslated region).

30

4) Poly-A tail: Added after nucleotide 2106 (as in Sequence I.D. No. 1) or approximately 150-200 bases downstream (precise site not determined): this variation probably originates from an alternative polyadenylation
35 signal.

5) Nucleotide 1214 = T (as in Sequence I.D. No. 1) or

C (as in Sequence I. D. No. 27), therefore codon 358 is either ATC or ACC, coding for isoleucine or threonine, respectively. This sequence variation is described in more detail in relation to Example II.

5 Analysis of the predicted MED1 protein sequence reveals a tripartite structure. At the amino terminus, MED1 contains a region of homology to the methyl-CpG binding domain (MBD) of MeCP2, a chromosomal protein which binds CpG-methylated DNA and may mediate the
10 effects of DNA methylation on chromatin structure and transcription (Lewis et al., (1990) *Cell* 69:905-914; Nan et al., (1993) *Nucleic Acids Res.* 21:4886-4892). The same region of MED1 is also homologous to the MBD of the human protein PCML, a component of the transcriptional
15 repressor MeCP1 (Cross et al., (1997) *Nat. Genet.* 16:256-259). The central portion of MED1 does not display a recognizable domain structure, but it appears to be rich in positively-charged amino acids, often arranged in short clusters which might represent nuclear
20 localization signals (Boulikas, T., (1993) *Critical Rev. in Eukaryotic Gene Expression* 3:193-227). Finally, at the carboxy terminus, MED1 contains a putative catalytic domain sharing homology with several bacterial
25 endonucleases involved in DNA repair, including MutY and endonuclease III from *E. coli*, ultraviolet endonuclease from *Micrococcus luteus*, and the putative endonuclease encoded by the ORF10 of the thermophilic archaeon *Methanobacterium thermoformicum*. See Figure 4A, 4B and
30 4C. A schematic of the domain organization of MED1 is shown in Figure 5.

 In order to confirm that the MED1 open reading frame is capable of directing the synthesis of a protein product, a construct of MED1 in the vector pCDNA3 was employed in an *in vitro* coupled transcription and
35 translation assay. The result indicated that the MED1 open reading frame drives the translation of two polypeptides of 70 and 65 kD, shown in Figure 6, in good

agreement with the molecular weight predicted from the amino acid sequence. The synthesis of these two polypeptides might be the result of initiation from the two close ATG codons, at nucleotide position 142 and 262, respectively. Such a possibility is known to occur as a result of "leaky" ribosome scanning and is increased by a suboptimal Kozak's context (Kozak, M., (1995) *Proc. Natl. Acad. Sci.* 92:2662-2666). The difference in molecular weight (5kD) would be compatible with the distance between the two ATG codons (40 a.a.).

To determine if two MED1 proteins are also synthesized *in vivo*, a hemagglutinin epitope was fused at the carboxyterminal end of the MED1 open reading frame, generating the construct MED1-HT. Constructs were also generated which fused a hemagglutinin tag immediately before each of the two putative initiation codons (HT-MED1-M1 and HT-MED1-M2). These constructs were transiently transfected in NIH-3T3 cells and lysates of the transfectants were prepared and separated by SDS-PAGE. Western analysis with an anti-hemagglutinin tag antibody revealed the presence of a band of approximately 72 kD in cells transfected with the carboxyterminally tagged MED1-HT. This band comigrates with the one present in HT-MED1-M1 transfectants, indicating that the first ATG at nucleotide position 142 is the initiation codon *in vivo*. See Figure 7. Since the expression level of HT-MED1-M1 which uses the hemagglutinin tag ATG codon is much higher than MED1-HT which uses the autologous ATG codon, it is possible that the expression of the MED1 protein is under a tight translational control.

Finally, the MED1 gene was mapped with fluorescence *in situ* hybridization to human chromosome 3q21. See Figure 8.

In order to determine whether MED1 has endonuclease activity, the catalytic (endonuclease) domain was expressed in *E. coli* as a carboxyterminal fusion to a

6xHis tag. High levels of expression of the domain as a polypeptide of approximately 18-22 kD were achieved. See Figure 9A, left panel. Bacterial lysates expressing the catalytic domain were separated in an activity polyacrylamide gel containing denatured calf thymus DNA. Following electrophoresis, the gel was incubated in a Tris-buffered solution containing 25% isopropanol and then in Tris buffer alone to allow digestion of DNA. DNA was visualized by staining the gel with toluidine blue O. Results revealed a zone of clearing, indicative of DNA digestion, migrating at approximately 18-22 kD in *E. coli* lysates expressing the endonuclease domain but not in control lysates. See Figure 9A, right panel. This experiment indicates that the recombinant endonuclease domain of MED1 displays deoxyribonuclease activity.

To better define its nuclease properties, the entire MED1 protein was expressed in *E. coli* as a carboxyterminal fusion to a six-histidine tag and purified on a nickel-agarose column to approximately 95% homogeneity. See Figure 9B, left panel. Endonuclease activity was assayed by evaluating the conversion of a supercoiled plasmid into open circles (nicked) and linear molecules. Increasing amounts of the purified MED1 protein were incubated with supercoiled plasmid DNA at 37°C for 30 min, and the products of the reactions, separated on a 1% agarose gel, were visualized by ethidium bromide staining. Incubation with MED1 resulted in a dose-dependent appearance of nicked and linearized molecules (Fig. 9B, right panel). In order to rule out the possibility that a bacterial endonuclease activity copurifying with MED1 is responsible for the observed effects, a deletion mutant lacking the putative endonuclease domain was also purified. This mutant failed to produce nicked and linearized DNA molecules (Fig. 9B, right panel). These results indicate that MED1 has single- and double-strand

endonuclease activity. Digestion of the MED1-linearized plasmid with the restriction enzyme EcoRI, which performs two closely spaced cuts on this plasmid, resulted in the appearance of a smear, indicating that
5 MED1 does not have preferential cutting sites on this substrate. The production of linear molecules by MED1 in the above assay is intriguing. The kinetics suggest rapid counter-nicking of the second strand across from a site where the first nick is formed. It will be
10 interesting to determine whether the MED1 nicks occur in CpG-rich regions and whether cytosine methylation inhibits the second nicking event.

To assess whether the MED1 methyl-CpG binding domain (MBD) is able to bind methylated DNA, a FLAG epitope was fused at the amino terminal end of the MED1
15 open reading frame, generating the construct FT-MED1/f5, and this construct was transfected into the human kidney line 293. Cells were also transfected with the empty expression vector. Seventy-two hours after
20 transfection, cells were lysed and the lysates were immunoprecipitated with an anti-Flag antibody coupled to agarose beads. Bound protein was eluted from the beads following incubation with a FLAG peptide. The FT-MED1/f5 and control eluates were incubated with a ³²P-
25 labeled double-stranded oligonucleotide containing a total of five fully methylated CpG sites, in the presence or absence of a 100-fold excess of the unlabeled or "cold" oligonucleotide. The binding reactions were separated on a non-denaturing
30 polyacrylamide gel and detected by autoradiography of the dried gel. A slowly migrating band was detected in the FT-MED1/f5 eluate lanes, but not in the control lane. This band was abolished by competition with excess cold oligonucleotide. This experiment indicated
35 that the MBD of MED1 functions as a specific methylated DNA binding domain in vivo. See Figure 10A.

To further characterize the DNA binding properties

of MED1, its putative methyl-CpG binding domain (MBD) was expressed in *E. coli* as a carboxyterminal fusion to a six-histidine tag, and it was purified by metal-chelating affinity chromatography followed by ion-exchange chromatography on SP Sepharose (Pharmacia). The purity of the MED1 MBD was estimated at >98% by SDS-PAGE followed by Coomassie staining. The purified MBD was incubated with a ³²P-labeled double-strand oligonucleotide of arbitrary sequence containing five symmetrical methyl-CpG sites. As a control, MBD was incubated with a ³²P-labeled double-strand oligonucleotide of identical sequence in which cytosines replaced methyl-cytosines. EMSA analysis of the complexes indicated that the MED1 MBD binds to methylated DNA and fails to bind to unmethylated DNA (Fig. 10B, lanes 2 and 6). Binding to the methylated probe was competed by preincubation with a 100-fold excess of cold methylated oligonucleotide (lane 3). Little competition was observed following preincubation with the unmethylated oligonucleotide (Fig. 10B, lane 4). This experiment provides further evidence of the methyl-CpG binding specificity of the MED1 MB

The physical association of MED1 with other DNA repair proteins was assessed as follows. 293 cells were transfected with the construct FT-MED1/f5 or with an empty expression vector. Seventy-two hours after transfection, cell lysates were prepared and immunoprecipitations carried out with anti-FLAG antibodies coupled to agarose beads. Immunoprecipitated proteins were separated by SDS-PAGE, transferred to membrane and probed with anti-hMSH2 antibody. The antibody detected a band of approximately 103 kD comigrating with hMSH2 in the anti-FLAG immunoprecipitate from FT-MED1/f5 transfected 293 cells but not from control cells. See Figures 11A and 11B. This experiment demonstrates the physical association of MED1 in a complex with hMSH2.

In order to confirm that the MLH1 / MED1 interaction detected in yeast also occurs in human cells, co-immunoprecipitation experiments were performed. Human kidney HEK-293 cells were transfected with a hemagglutinin-tagged construct of MED1 (HT-MED1) or with an empty expression vector. Seventy-two hours after transfection, cell lysates were prepared and immunoprecipitations were carried out with an antibody directed against the hemagglutinin tag. Immunoprecipitated proteins were separated by SDS-PAGE, transferred to a membrane and probed with an anti-MLH1 monoclonal antibody. The antibody detected a band of approximately 82 kD co-migrating with MLH1 in the anti-hemagglutinin immunoprecipitate from HT-MED1-transfected HEK-293 cells but not from control cells (Fig. 11C). This experiment suggests that MED1 is present in a complex with MLH1.

EXAMPLE II

Identification of Mutations in MED1 in HNPCC patients

Mutational screening of the MED1 gene has been performed in ten HNPCC patients. Earlier studies on these patients revealed that they were negative for hMSH2 and hMLH1 mutations (Viel et al., (1997) Genes Chromosom Cancer 18:8-18). Polymerase chain reaction (PCR) amplification of MED1 fragments with MED1-specific primer oligonucleotides (provided in Table I), has been performed followed by direct sequencing of PCR products. A sequence variant which converts isoleucine 358 to threonine (I358T) has been identified in the germ-line of a female patient affected by two independent synchronous colon cancers. Analysis of one of the cancers revealed the loss of a normal allele. This finding is in agreement with a possible tumor suppressor role of MED1. The I358T variant is presently being searched in other affected and unaffected individuals of the family to determine if it cosegregates with the disease. Thus, the I358T variant is present at a

frequency of 1 out of 10 HNPCC patients (10%). This variant is also present in the general population at a lower frequency of approximately 3 out of 69 individuals (4.3%). Taken together these findings suggest that the
5 I358T variant of MED1 may be associated with an increased risk for colon cancer.

EXAMPLE III

Screening Cancer Patient DNA Samples 10 for Mutations in MED1

A panel of 14 sporadic colorectal cancers with microsatellite instability but with no detectable defect in the two major mismatch repair genes, hMSH2 and hMLH1 (Y. Wu et al *Genes Chromosomes and Cancer* 18, 269: 1997)
15 were screened for mutations by PCR amplification of all the MED1 exons from genomic DNA, followed by direct sequencing of PCR products with an automated DNA sequencer (ABI), using the primers shown in Table I. Sequence analysis revealed MED1 mutations in 4 of 14
20 (28.6%) tumors. In all four of these tumors, a one-base deletion occurred in one of two mononucleotide repeats [(A)6 and (A)10] located in the coding region of MED1 (Fig. 13A and 13B) (Mutations were confirmed by sequencing at least three independent PCR products on
25 both strands); the mutations were somatic, as they were not detected in the corresponding peripheral blood DNA. The one-base deletions cause frameshifts and predict the synthesis of truncated proteins (Fig. 13C). These alterations resemble the frameshift mutations described
30 in the (A)8 and (C)8 tracks present in the coding region of the mismatch repair genes MSH3 and MSH6, respectively (S. Malkhosyan et al *Nature* 382 499 :1996). Furthermore, these alterations appear to be selected for in tumor cells, as similar (A)n mononucleotide repeats,
35 including the (A)8 stretch in the coding region of PMS2, are not altered in this tumor panel. Similarly, preliminary screening experiments of 26 endometrial

cancer patients led to the identification of a mutation in MED1.

TABLE II

	Patient	Sex	Tumor Site	Age Diagnosis	MED1 Mutation	Codon	Result
5	c18T	F	caecum	83	(A)10 to (A)9	310-313	frameshift and stop at codon 317
10	C220T	M	transverse colon	79	(A)10 to (A)9	310-313	same as above
15	c226T	F	ascending colon	70	(A)10 to (A)9	310-313	same as above
20	c215T	F	caecum	66	(A)6 to (A)5	280-282	frameshift and stop at codon 317
25	UPN252T	F	endometrium	N/A	(A)10 to (A)9	310-313	frameshift and stop at codon 317

Discussion

Two long-standing and closely related issues in eukaryotic mismatch DNA repair are identifying the endonuclease activity responsible for incising the DNA strand carrying the mutation, and defining the nature of the strand-targeting signal. In *E. coli*, Muth performs this function through the recognition of hemimethylated d(GATC) sites. However, eukaryotic functional homologues of Muth are not currently known. Due to the lack of information on the molecular determinants of strandedness, it was hypothesized that a reasonable approach towards the cloning of eukaryotic Muth functional homologues would be to identify hMLH1 interactors. By analogy with the MutL-Muth interaction in the bacterial system, the eukaryotic mismatch repair endonuclease is expected to be a hMLH1 interactor.

Accordingly, the "interaction cloning" of MED1, a gene encoding a viable candidate for the mismatch repair endonuclease is described in the previous examples. The MED1 protein has several features compatible with such a role. MED1 specifically interacts with hMLH1 in the yeast system and mammalian cells, and with hMSH2 in a mammalian

cell system. Whether MED1 interacts with other components of the mismatch repair complex, such as hMSH3, hMSH6/GTBP and hPMS2 has yet to be determined. MED1 has a catalytic domain showing homology to several bacterial DNA repair endonucleases, and it is predicted that MED1 would have N-glycosylase and possibly apurinic or apyrimidinic (AP) lyase activities. Among the MED1 homologues, both the *E. coli* MutY and endonuclease III, and the *M. luteus* UV-repair endonuclease have DNA N-glycosylase and AP endonuclease activities. Interestingly, MutY is active on A.C, A.G and A.8-oxoG mismatches, whereas endonuclease III is active on mismatches containing some damaged derivatives of thymidine and cytosine. The homology between MED1 and the ORF10-encoded protein of *M. thermoformicicum* (Nolling et al., (1992) *Nucleic Acids Res.* 20:6501-6507) is particularly intriguing. It has been proposed that this open reading frame encodes a mismatch DNA repair enzyme, functionally associated with the methylase of the *M. thermoformicicum* restriction/modification system. ORF10 would be active on G/T mismatches originated by deamination of 5-methyl-cytosine, a product of the methylase, to thymidine under thermophilic conditions. Spontaneous deamination of 5-methyl-cytosine in CpG dinucleotides to thymidine (G.m5C → G.T) is a source of endogenous mutations in the human genome (Rideout et al., (1990) *Science* 249:1288-1290). Almost 50% of the p53 point mutations in colorectal cancer are transitions at CpG dinucleotides (Greenblatt et al., (1994) *Cancer Res.* 54:4855-4878). Conservation of MED1-related sequences involved in mismatch repair in organisms belonging to two distant phyla (Eubacteria and Archeobacteria) suggests that human MED1 is an endonuclease active on DNA mispairs.

A common feature of the MED1-related endonucleases is the presence of a Cys-X6-Cys-X2-Cys-X5-Cys sequence at their carboxy terminus. This sequence, as shown in endonuclease III, ligates the [4Fe-4S] iron-sulfur cluster and defines a novel DNA binding motif (named the FCL

motif), which provides the correct alignment of the enzyme along the DNA (Thayer et al., (1995) *Embo J.* 14:4108-4120). MED1 lacks a FCL motif at its carboxy terminus, but contains a methyl-CpG DNA binding domain at the amino 5 terminus.

The presence of this methyl-CpG binding domain in MED1 suggests a mechanism for strand-determination. In human mismatch repair, strand-specificity may be determined by the MED1-mediated recognition of methyl-CpG sequences. The 10 newly synthesized strand would be recognized as such by virtue of its transient lack of CpG methylation after replication as shown in Figure 12. In this model, cytosine methylation in eukaryotes would be functionally equivalent to adenine methylation in *E. coli*, as is the case for 15 methylation-mediated transcriptional repression. This model is consistent with experimental evidence suggesting that, in monkey CV1 cells, cytosine hemimethylation at CpG sites may be a determinant of strandedness (Hare et al., (1985) *Proc. Natl. Acad. Sci.* 82:7350-7354). Since a nick 20 in one of the DNA strands is capable of efficiently directing the mismatch repair *in vitro*, it is also possible that DNA termini generated at the replication fork represent the only strand-targeting signal in organisms lacking genome methylation such as *Drosophila* and *S. cerevisiae*. Accordingly, screening of the *S. cerevisiae* 25 genome database did not identify any homologue of MED1. However, in CV1 cells, single-strand nicks were shown to synergize with CpG hemimethylation in directing repair, indicating that multiple mechanisms may play a role in 30 strand determination. Thus, our data would imply that epigenetic modification of the genome via cytosine methylation not only participates in X-chromosome inactivation, imprinting, and transcriptional repression (A. Bird, *Cell* 70: 5-8, 1992; R. A. Martienssen and E. J. 35 Richards, *Curr. Opin. Genet. Dev.* 5: 234-242, 1995), but is also involved in DNA repair. Indeed, recent studies propose that DNA methylation plays a role in maintaining

genomic stability (C. Lengauer, K. W. Kinzler, B. Vogelstein, *Proc. Natl. Acad. Sci USA* 94: 2545-2550, 1997)).

The interpretation of the MED1 mutational data requires some caution. Although it is presently unclear whether MED1 mutations promote or are the consequence of microsatellite instability, their apparent selection in tumors suggest that they may contribute to the unfolding of tumor genomic instability, as has been proposed for the MSH3 and MSH6 coding microsatellite mutations (M. Perucho, *Nature Med* 2: 630-631, 1996). Due to the variable amount of contaminating normal cells in primary tumor specimens, it is difficult to determine the homozygous or heterozygous nature of the MED1 mutations. Sequence analysis (Fig. 13) shows apparent retention in the tumors of the wild-type MED1 allele. This may indicate that the products of the mutant alleles, which lack the endonuclease domain (Fig. 13C), act in a dominant negative fashion, perhaps competing for methyl-CpG DNA binding. Alternatively, the heterozygous mutations may reduce the total amount of functional molecules (haploinsufficiency).

In summary, although the endonuclease domain of MED1 does not display a significant homology to Muth, the specific interaction with hMLH1 and the domain organization indicate that MED1 may be a functional homologue of Muth, i.e. the a DNA repair endonuclease capable of strand discrimination. Assuming MED1 is the long-sought eukaryotic homologue of muth, then, like other mismatch repair genes which are mutated in HNPCC as well as in sporadic cancers with microsatellite instability, MED1 is a candidate gene for cancer genetic testing, both in HNPCC families and in sporadic cancers with microsatellite instability. It should be noted that only about 70% of HNPCC cases and only about 65% of sporadic tumors with microsatellite instability carry mutations in the known mismatch repair genes hMSH2, hMLH1, hPMS2 and hPMS1. The remainder 30-35% of the cases have an as yet unidentified

mismatch repair defect and a fraction may therefore harbor mutations or loss of expression of MED1. Indeed, frameshift MED1 mutations were detected in both colorectal and endometrial cancers. See Figure 13 and Table II.

5 While certain preferred embodiments of the present invention have been described and specifically exemplified above, it is not intended that the invention be limited to such embodiments. Various modifications may be made to the invention without departing from the scope and spirit
10 thereof as set forth in the following claims.

What is claimed is:

1. An isolated double-stranded nucleic acid molecule which upon denaturation, specifically hybridizes with SEQ ID NO: 1, said nucleic acid molecule comprising a
5 sequence encoding a human endonuclease about 580 amino acids in length, said encoded endonuclease comprising an amino-terminal methyl CpG-binding domain, an internal segment rich in positively charged amino acids and a carboxy-terminal catalytic domain, said catalytic domain
10 having deoxyribonuclease activity.
2. The nucleic acid molecule of claim 1, which is DNA.
- 15 3. The DNA molecule of claim 2, which is a cDNA comprising a sequence approximately 2.4 kilobase pairs in length that encodes said human endonuclease.
4. The DNA molecule of claim 2, which is a gene
20 comprising introns and exons, the exons of said gene specifically hybridizing with the nucleic acid of SEQ ID NO: 1, and said exons encoding said human endonuclease protein.
- 25 5. The nucleic acid molecule of claim 1, which is RNA.
6. A vector comprising the nucleic acid molecule of claim 1.
- 30 7. A host cell comprising the vector of claim 6.
8. The nucleic acid molecule of claim 1, wherein said nucleic acid encodes a human endonuclease protein
35 comprising an amino acid sequence selected from the group consisting of an amino acid sequence encoded by SEQ ID NO: 2 and natural allelic variants of said nucleic acid.

9. The nucleic acid molecule of claim 8, which comprises SEQ ID NO: 1.

10. An isolated nucleic acid molecule comprising
5 a sequence selected from the group consisting of:

- a) SEQ ID NO: 1;
- b) a sequence which specifically hybridizes with SEQ ID NO: 1;
- c) a sequence encoding a polypeptide of SEQ
10 ID NO: 2; and
- d) a nucleic acid sequence encoding a catalytic domain of an endonuclease protein having an amino acid sequence corresponding to amino acids 455-580 of SEQ ID NO: 2.

15

11. An oligonucleotide between about 10 and about 200 nucleotides in length, which specifically hybridizes with a nucleotide sequence encoding amino acids of SEQ ID NO: 2.

20

12. An oligonucleotide between about 10 and about 200 nucleotides in length, which specifically hybridizes with a sequence in the nucleic acid molecule of claim 1, said sequence encoding the methyl CpG binding
25 domain of said endonuclease protein.

13. An isolated human endonuclease protein, about 580 amino acids in length, said encoded protein comprising an amino-terminal methyl CpG-binding domain, an
30 internal segment rich in positively charged amino acids and a carboxy-terminal catalytic domain, said catalytic domain having deoxyribonuclease activity.

14. An antibody immunologically specific for the
35 isolated protein of claim 13.

15. An antibody as claimed in claim 14, said

antibody being monoclonal.

16. An antibody as claimed in claim 14, said antibody being polyclonal.

5

17. A pharmaceutical composition comprising a polypeptide as claimed in claim 13 and a pharmaceutically acceptable carrier.

10 18. A pharmaceutical composition comprising an antibody as claimed in claim 14 and a pharmaceutically acceptable carrier.

19. A method of diagnosing a susceptibility or
15 predisposition to cancer in a patient caused by an alteration in a MED1 encoding nucleic acid, wherein said patient sample is analyzed by a method selected from the group consisting of:

a) a method of comparing a sequence of nucleic
20 acid in the sample with the MED1 nucleic acid sequence to determine whether the sample from the patient contains mutations; and

b) a method of determining the presence, in a sample from a patient, of a polypeptide encoded by the MED1
25 nucleic acid and, if present, determining whether the polypeptide is altered; and

c) a method of DNA restriction mapping to compare the restriction pattern produced when a restriction enzyme cuts a sample of nucleic acid from the patient with
30 the restriction pattern obtained from normal MED1 gene or from known mutations thereof; and

d) a method employing a specific binding member capable of binding to a MED1 nucleic acid sequence, the specific binding member comprising nucleic acid
35 hybridizable with the MED1 sequence; and

e) a method wherein at least one antibody domain with specificity for an epitope selected from the group

consisting of a native MED1 nucleic acid sequence epitope, or a polypeptide epitope, the specific binding member being labelled so that binding of the specific binding member to its binding partner is detectable; and

5 f) a method of PCR amplification involving one or more primers based on normal and mutated MED1 gene sequence to screen for normal and mutant MED1 gene in a sample from a patient.

10 20. A method of identifying a target nucleic acid molecule in a test sample using a nucleic acid probe having the sequence shown in SEQ ID NO: 1, the method comprising contacting the probe and the test sample under hybridizing conditions and observing whether hybridization takes place.

15

21. A method according to claim 20 wherein the probe is used to identify a nucleic acid selected from the group consisting of a MED1 nucleic acid sequence and a mutant allele thereof.

20

22. A kit for detecting mutations in a MED1 gene associated with a susceptibility to cancer, the kit comprising at least one nucleic acid probe(s) capable of specifically binding a mutated MED1 nucleic acid.

25

23. A kit for detecting mutations in a MED1 gene associated with susceptibility to cancer, the kit comprising at least one antibody capable of specifically binding a polypeptide encoded by a mutated MED1 nucleic acid sequence.

30

24. A kit comprising a pair of oligonucleotide primers having sequences corresponding to a portion of a nucleic acid sequence set out in SEQ ID NO: 1 for use in
35 amplifying a nucleic acid selected from the group consisting of a MED1 nucleic acid sequence and a mutant allele thereof.

25. A kit for determining the presence of at least one mutation in a sample of nucleic acid from an individual, the kit comprising:

a) a solid support having immobilized thereon at least 5 one allelic variant specific nucleic acid probes having sequences corresponding to portions of the sequence set out in SEQ ID NO:1 capable of specifically binding a mutated MED1 nucleic acid sequence; and

b) a detectable label for marking the presence of 10 sample nucleic acid hybridized to the probe(s).

26. A kit for determining the presence of at least one mutation in a sample of nucleic acid from an individual, the kit comprising:

15 a) a solid support having immobilized thereon at least one antibody capable of specifically binding a polypeptide encoded by a mutated MED1 nucleic acid sequence; and b)

b) a detectable label for marking the presence of antibodies bound to the sample polypeptides.

20

27. A method of screening for substances which modulate the activity of a MED1 polypeptide, the method comprising contacting at least one test substance with the MED1 polypeptide in a reaction medium, testing the activity 25 of the treated MED1 polypeptide and comparing that activity with the activity of native, untreated MED1 polypeptide in a comparable reaction medium.

28. A method as claimed in claim 27, wherein said 30 test substance is a mimetic of the MED1 polypeptide.

29. A chimeric animal comprising an exogenous MED1 allele.

LexA-hMLH1 / B42-f5

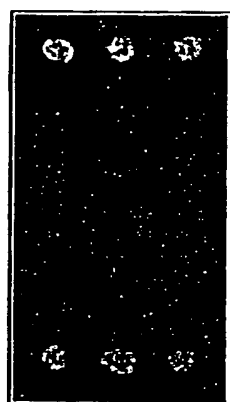
LexA / B42-f5

LexA-myc / B42-f5

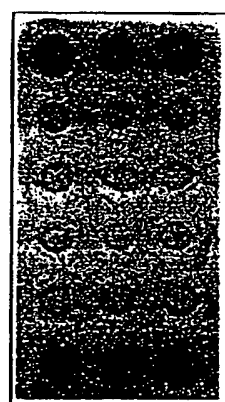
LexA-bicoid / B42-f5

LexA-K-rev1 / B42-f5

LexA-K-rev-1 / Krit1



Leu-



X-gal

FIGURE 1

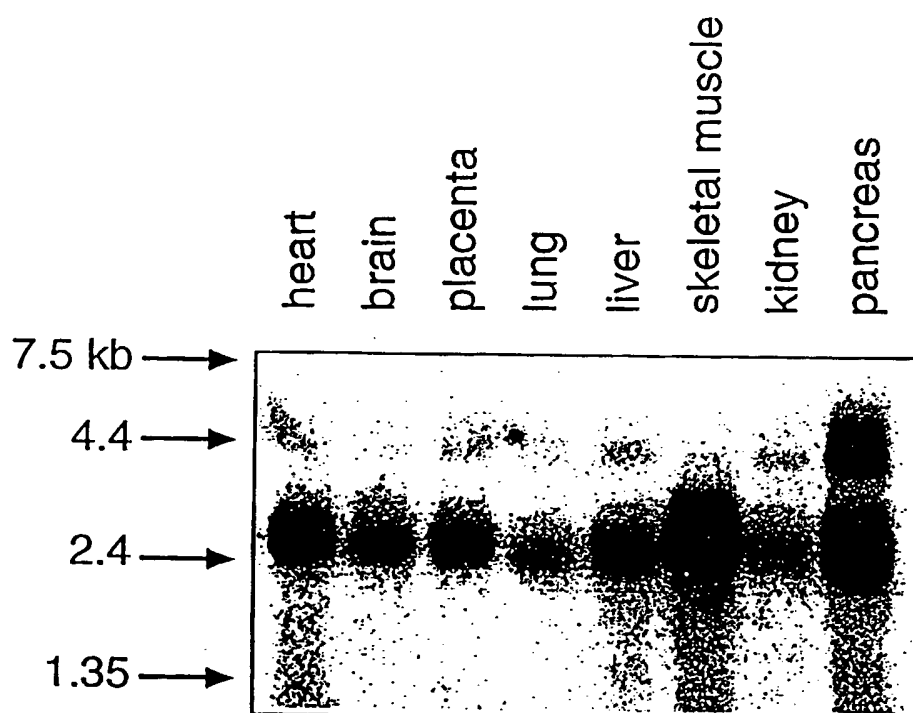


FIGURE 2

1 GGCGGCGTCTGGGGCGCTTTTCGCAACATTCAGACCTCGGTTGCAGCCCGGTGCGGTGAGCTGAAGAGGTTTCACATCTTACTCCGCCCCA 90
91 CACCTTGGGCGTTCGGCGCTGGGCTCGTTGCTGCAGCCGACCTGCTCGATGGGCAAGGCTGGGCTGGAGAGTCTGAGTCTGGGGGAC 180
M G T T G L E S L S L G D
181 CGCGGAGCTGCCCCACCGTCACCTCTAGTGAGCGCCTAGTCCCAGACCCGCGAATGACCTCCGCAAGAAGATGTTGCTATGGAATTG 270
R G A A P T V T S S E R L V P D P P N D L R K E D V A M E L
271 GAAAGATGGGAGAAGATGAGGAACAAATGATGATAAAAAGAAGCAGTGAATGTAATCCCTTGCTACAAGAACCCATCGCTTCTGCTCAG 360
E R V G E D E E Q M M I K R S S E C N P L L Q E P I A S A Q
361 TTTGGTGTCTACTGCAGGAACAGAAATGCCGTAAGTCTGTCCCATGTGGATGGGAAAGAGTTGTGAAGCAAAGGTTATTTGGGAAGACAGCA 450
F G A T A G T E C R K S V P C G W E R V V K Q R L F G K T A
451 GGAAGATTTGATGTGTACTTTATCAGCCCAAGGACTGAAGTTCAGATCCAAAAGTTCACTTGCTAATTATCTTCACAAAATGGAGAG 540
G R F D V Y F I S P Q G L K F R S K S S L A N Y L H K N G E
541 ACTTCTCTTAAGCCAGAAGATTTTGATTTTACTGTACTTTCTAAAAGGGGTATCAAGTCAAGATATAAAGACTGCAGCATGGCAGCCCTG 630
T S L K P E D F D F T V L S K R G I K S R Y K D C S M A A L
631 ACATCCCATCTACAAAACCAAGTAACAATTCAACTGGAACTCAGGACCCGAAGCAAGTGCAAAAAGGATGTGTTTATGCCGCCAAGT 720
T S H L Q N Q S N N S N W N L R T R S K C K K D V F M P P S
721 AGTAGTTTCAGAGTTCAGGAGAGCAGAGGACTCTCTAACTTTACTTCCACTCATTGCTTTTGAAAGAAGATGAGGTTGTGATGATGTT 810
S S S E L Q E S R G L S N F T S T H L L L K E D E G V D D V
811 AACTTCAGAAAGGTTAGAAAGCCCAAGGAAAGGTGACTATTTTGAAAGGAATCCCAATTAAGAAAACATAAAAAGGATGTAGGAAGAGC 900
N F R K V R K P K G K V T I L K G I P I K K T K K G C R K S
901 TGTTCAGGTTTGTTCACAAAGTGATAGCAAAAGAGAATCTGTGTGTAATAAAGCAGATGCTGAAAGTGAACTGTTGCACAAAAAAGTCAG 990
C S G F V Q S D S K R E S V C N K A D A E S E P V A Q K S Q
991 CTTGATAGAAGTCTGCTGCAATTTCTGATGCTGGAGCATGTGGTGAGACCCCTCAGTGTGACCAAGTGAAGAAAACAGCCTTGTAAAAA 1080
L D R T V C I S D A G A C G E T L S V T S E E N S L V K K K
1081 GAAAGATCATTGAGTTTCAGGATCAAAATTTTGTCTGAACAAAAAATCTTGGCATCATAAACAATTTTGTTCAGCCAAAGACTCAGAA 1170
E R S L S S G S N F C S E Q K T S G I I N K F C S A K D S E
1171 CACAAAGAGAAGTATAGGATACCTTTTAGAATCTGAAGAAATCGGAACAAAAGTAGAAGTTGTGGAAAGGAAAGAACATTGTCATCT 1260
H N E K Y E D T F L E S E E I G T K V E V V E R K E H L H T
1261 GACATTTTAAACGTTGGCTCTGAAATGGACAACAACCTGCTACCAACCCAGGAAGACTTCACTGGTGAGAAAAATATTTCAAGAAGATACC 1350
D I L K R G S E H D N N C S P T R K D F T G E K I F Q E D T
1351 ATCCAGGAACACAGATAGAAAGAAGGAAACAAAGCCTGTATTTTCCAGCAATATAACAAAGAAGCTCTTAGCCCCCAGCAGCTAAA 1440
I P R T Q I E R R K T S L Y F S S K Y N K E A L S P P R R K
1441 GCCTTTAAGAAATGGACACCTCCTCGGTCACTTTTAACTCTGTTCAAGAAACACTTTTTCATGATCCATGGAAGCTTCTCATCGCTACT 1530
A F K K W T P P R S P F N L V Q E T L F H D P W K L L I A T
1531 ATATTTCTCAATCGGACCTCAGGCAAAATGGCAATACCTGTGCTTTGGAAGTTTCTGGAGAAGTATCCTTCAGCTGAGGTAGCAAGAACC 1620
I F L N R T S G K M A I P V L W K F L E K Y P S A E V A R T
1621 GCAGACTGGAGAGATGTGTAGAACTTCTTAAACCTCTTGGTCTCTACGATCTTCGGGCAAAAACCATTTGTCAAGTTCTCAGATGAATAC 1710
A D W R D V S E L L K P L G L Y D L R A K T I V K F S D E Y
1711 CTGACAAAGCAGTGGAAAGTATCCAATTGAGCTTTCATGGGATTGGTAAATATGGCAACGACTCTTACCGAATTTTGTGTCAATGAGTGG 1800
L T K Q W K Y P I E L H G I G K Y G N D S Y R I F C V N E W
1801 AAGCAGGTGCACCCCTGAAGACCACAAATTAATAAATATCATGACTGGCTTTGGGAAAAATCATGAAAAATTAAGTTTATCTTAAACTCTG 1890
K Q V H P E D H K L N K Y H D W L W E N H E K L S L S *
1891 CAGCTTTCAAGCTCATCTGTTATGTCATGCTTTGCACTTCAAAAAGCTTAATTAAGTACAACCAACCACTTTCCAGCCATAGAGATTT 1980
1981 TAATTAGCCCACTAGAAGCCTAGTGTGTGCTTTCTTAATGTGTGTGCCAATGGTGGATCTTTGCTACTGAATGTGTTTGAACATGTT 2070
2071 TTGAGATTTTAAAAATAAATTTATTTTGACAACAAAAA 2152

FIGURE 3

hmed1	KEDVAMELER	V.....	GEDEEQMMIK	RSSECNPLLQ	EPIAS
rmecp2	KEDKEGKHEP	LQPSAHHSAE	PAEAGKAETS	ESSGSAPAVP	EASAS
hmed1	AQ.....	FGA	PCGWERVVKQ	RLF GKTAGRF	DVYFI
rmecp2	PKQRRSIIIRD	RGPMYDDPTL	PEGWTRK LKO	RKSGRSAGKY	DVYLI
hmed1	SPQGLKFRSK	SSLANYLHKN	GETSLKPEDF	DFTVLSKRG I	KSR
rmecp2	NPOGKAFRSK	VELIAYFEKV	GDTSLDPNDF	DFTVTC . RGS	PSR

FIGURE 4A

a-hmedl	DPWKLLIA TI	IPVLWKEL EK	YPSAEVART A	DWRVSEEL K	PLGLYDDLRA K
b-endo3ecoli	SPPFELLIA VL	NKATAKEL YPV	ANTPAAML EL	GVSEVVKTY I K	TIGLYNSKAE
c-gtnrmt	DPYVLLITE I	KKIYDKKFFVK	YKCFEDILKT	PKSEVAKDI K	EIGLSNQRAE
d-uvendomi	TPPFELLIA TV	NAAATPAE FPAR	FPDAAHMAAA A	TEPSEVQELV R	STGSEYRNKAS
e-mutyecoli	TPYKXVWLESEV	IPYFEEFEMAR	FPTVTDLANA	PLDEVLHLWT	GLGY.YARAR
a-hmedl	TKQ	WKYPIELHG	WGRKYGND SY	NTAFGGWPTIA	RI F C VNE
b-endo3ecoli	EQHNGEVPED	RAALEA.LPG	VGRKKTANVVL	CLAFGGKKAAM	VDTHIFRV CN
c-gtnrmt	NDYGGGRVPRN	RKAILD.LPG	VGRKKTCAAVL	GNAFGGQPGIT	VDANFVRVIN
d-uvendomi	GRHGDGEVPAR	LEDLVA.LPG	VGRKKTAFVVL	SLSGKHPFI	VDTHFGR LAR
e-mutyecoli	TLHGKGKFPET	FEEVAA.LPG	VGRS TAGA	SLSGKHPFI	EDGNVKKRVLA
a-hmedl	LNKYYHDDWLME	NHEKLSLS	CHHW	HGRYTCIARK	PRCGS C I I E D
b-endo3ecoli	EQVBEKLLKV	VPA..EFKVD	CRDFNLGLMD	FSARII CAPRK	PKCEKCCGMSK
c-gtnrmt	LNYNHKA LWE	LAETLVPGGK	S.HR..LIF	HGARRVCHARR	PACCGCCPIAR
d-uvendomi	GKGRARRGPP	VPPARDWTML	VERFNQA	LGAMIC	PKCSLCP LQN
e-mutyecoli	KKEVENKLS	LS EQVT PAVG			
a-hmedl	LC EYK EKV DI				
b-endo3ecoli	LC SY Y E K C S T				
c-gtnrmt	WC PSY A A G E T				
d-uvendomi	GC I A A A N N S W				
e-mutyecoli					

FIGURE 4B

MED1 I A S A Q F G A T A G T E C R K S V P C G W E R V V K Q R L F G K T A G R F D V Y F I S P Q G L K F R S K S S L A N Y L 60
 PCM1 M A E D W L D C P A L G P G W K R R E V F R K S G A T C G R S D T Y Y Q S P T G D R I R S K V E L T R Y L 53

MED1 H K N G E T S L K P E D F D F T V L S K R G I K S 85
 PCM1 G P A C D L T . . . L F D E K Q G I L C Y P A P K A 76

FIGURE 4C

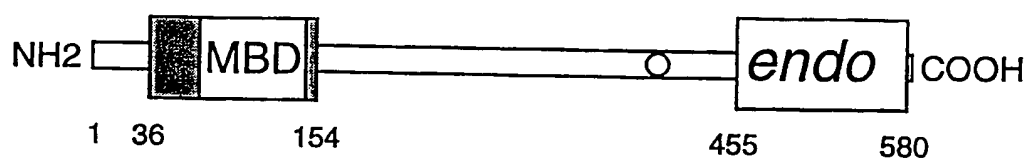
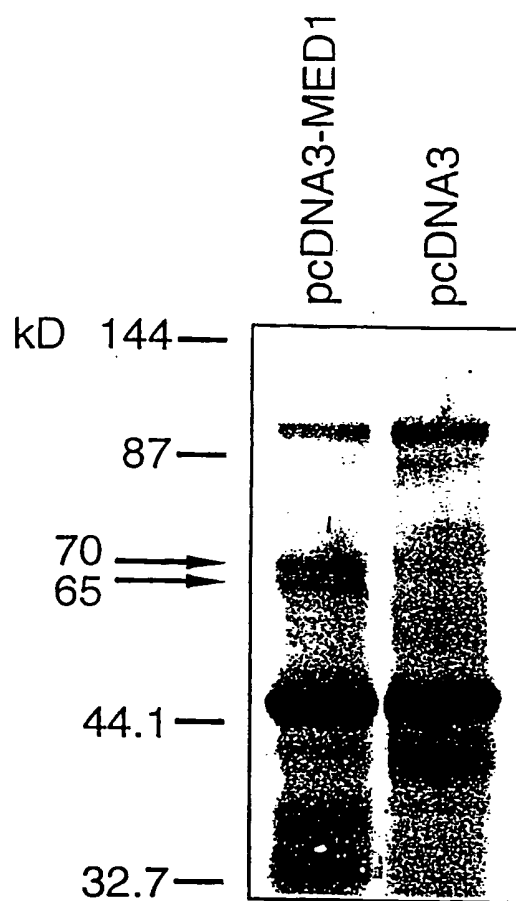
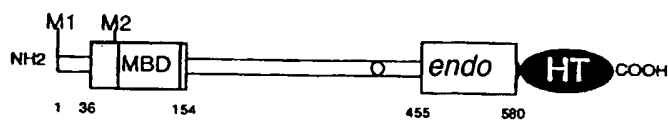


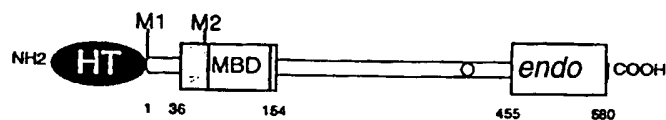
FIGURE 5

**FIGURE 6**

MED1-HT



HT-MED1-M1



HT-MED1-M2



FIGURE 7A

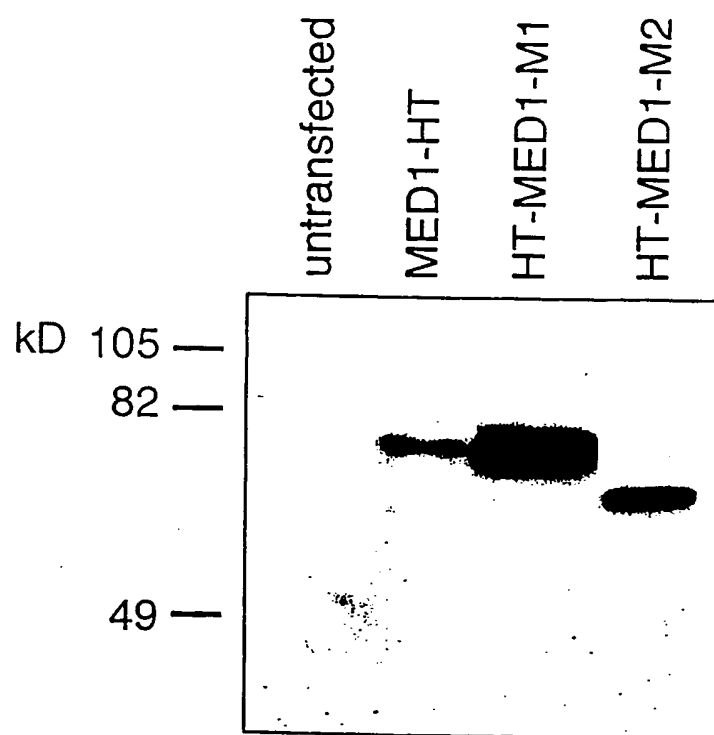


FIGURE 7B

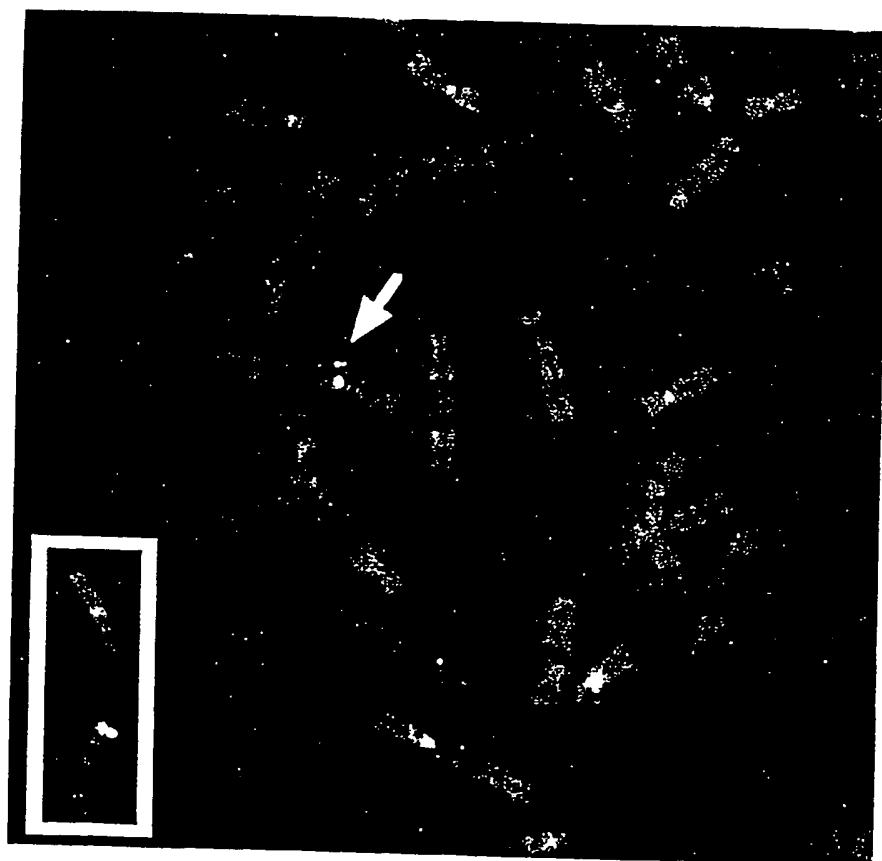
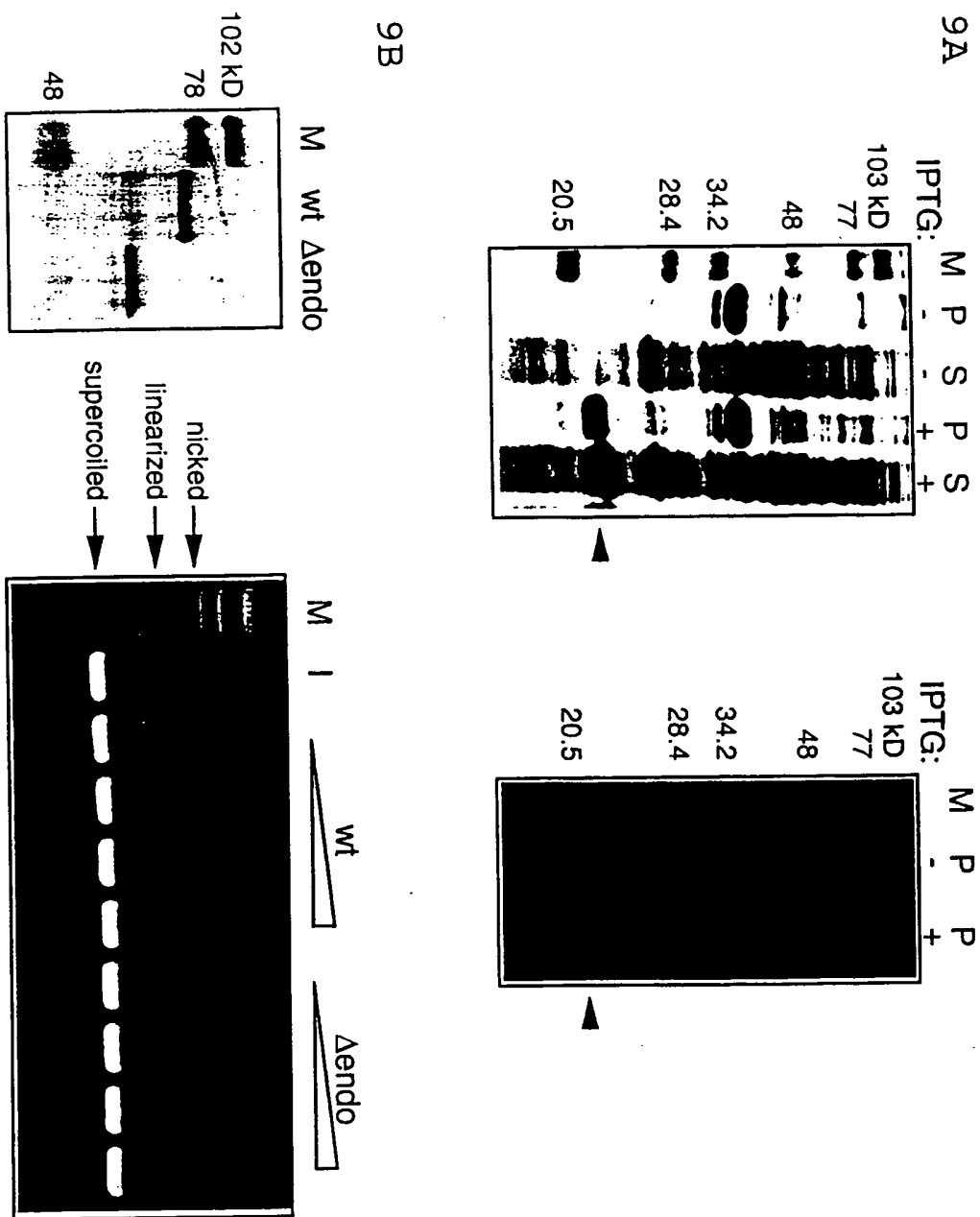


FIGURE 8



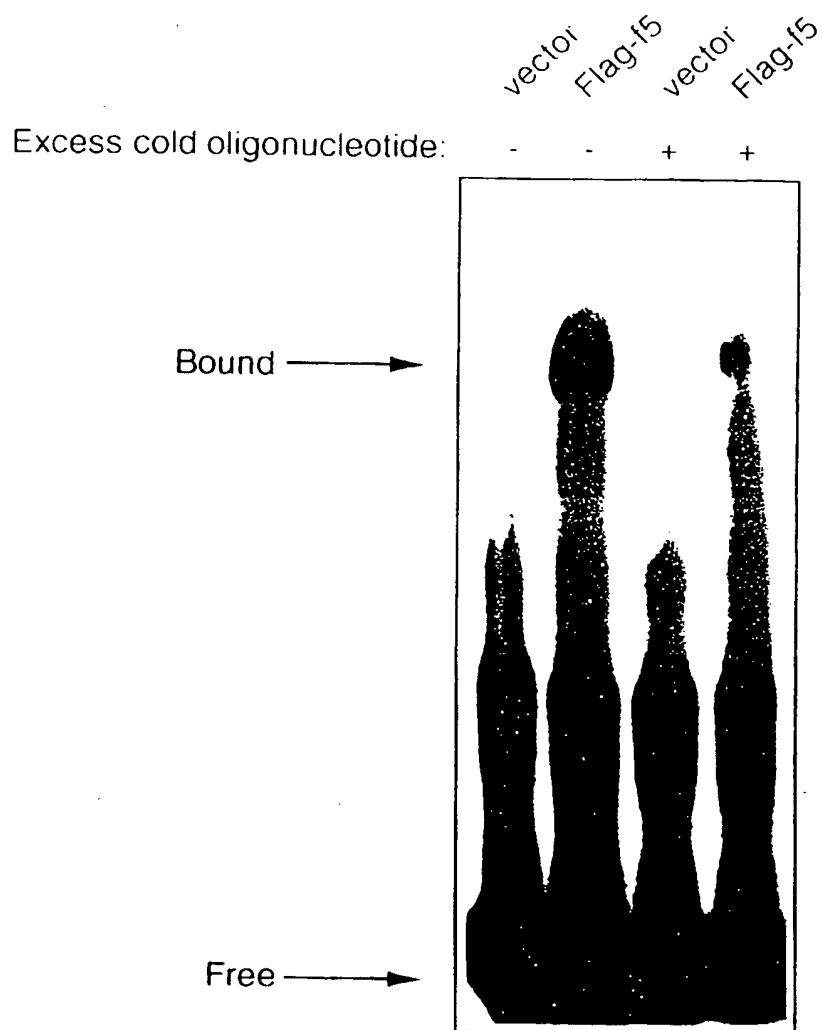


Fig. 10A

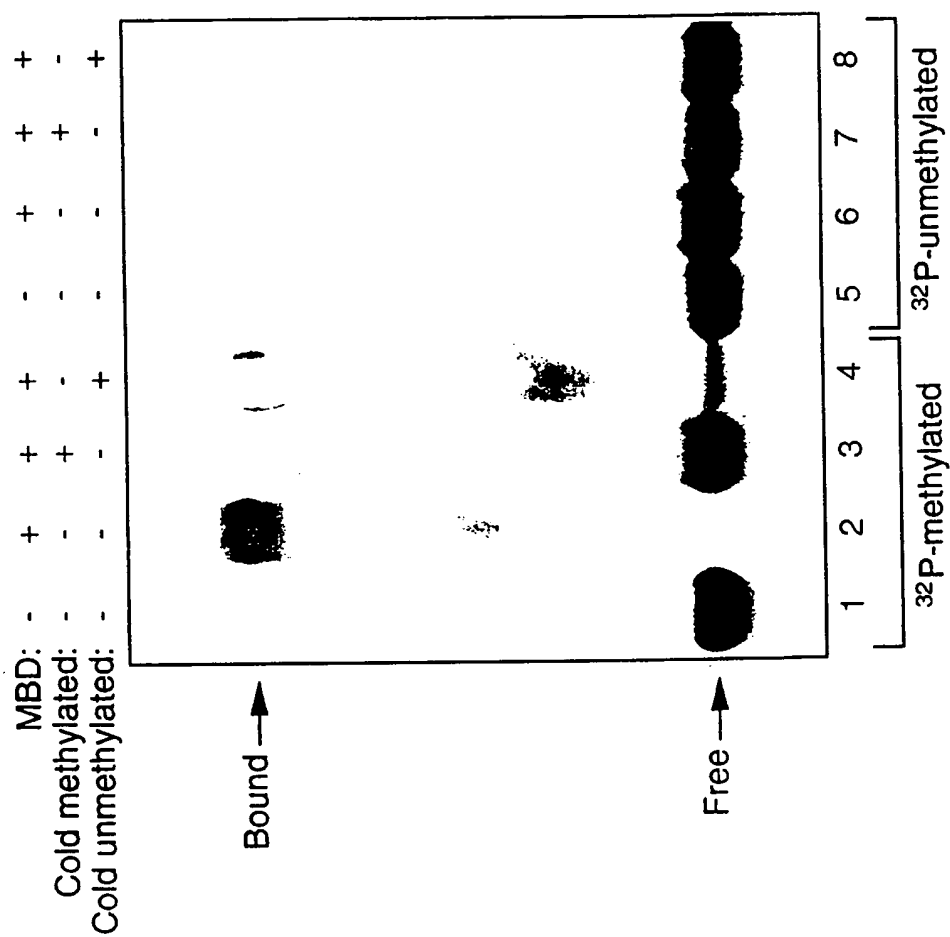


Fig. 10B

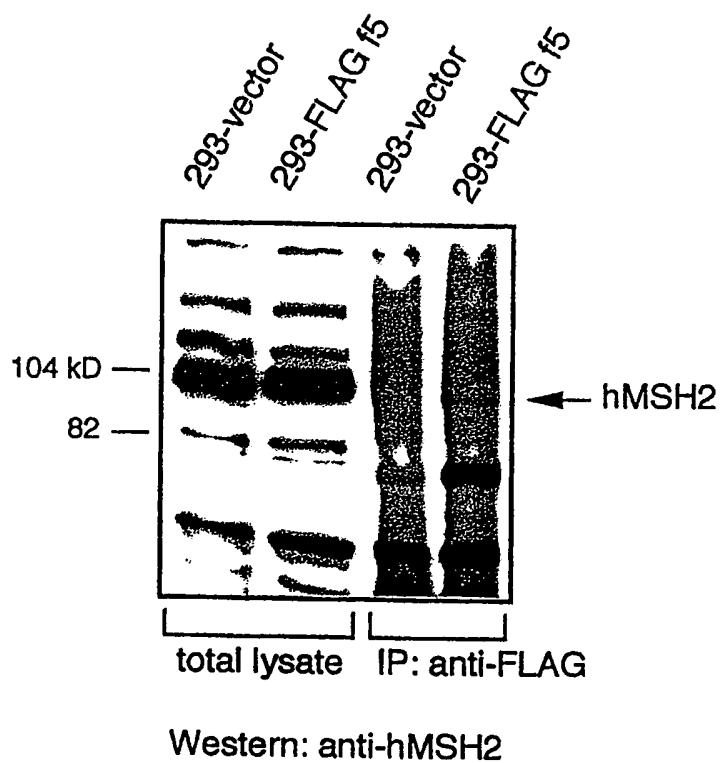
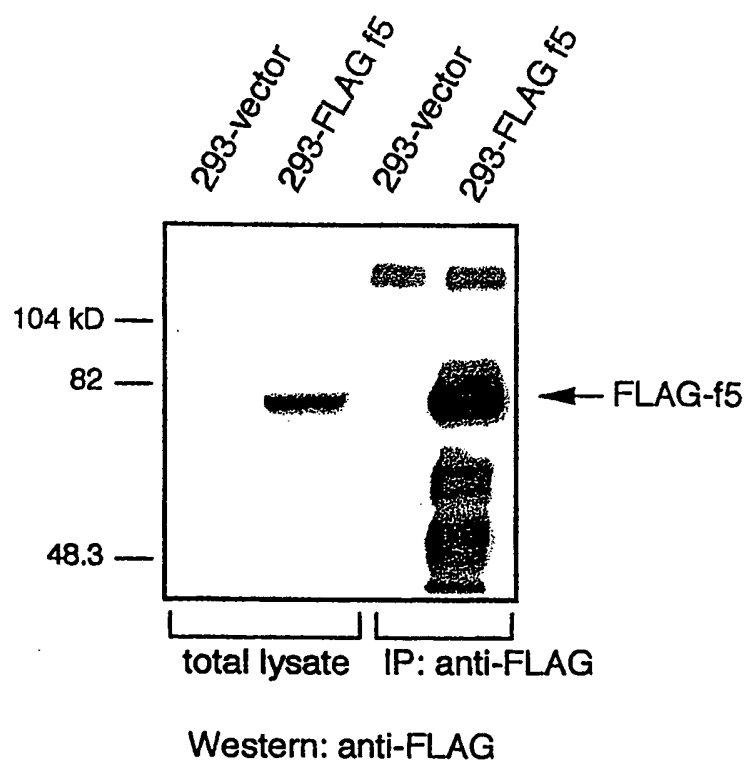


FIGURE 11A

**FIGURE 11B**

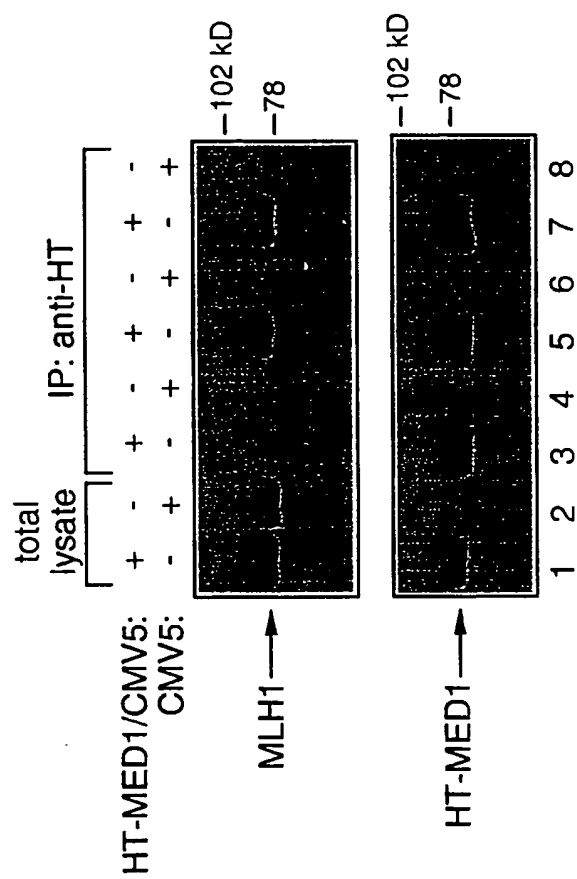


Fig. 11C

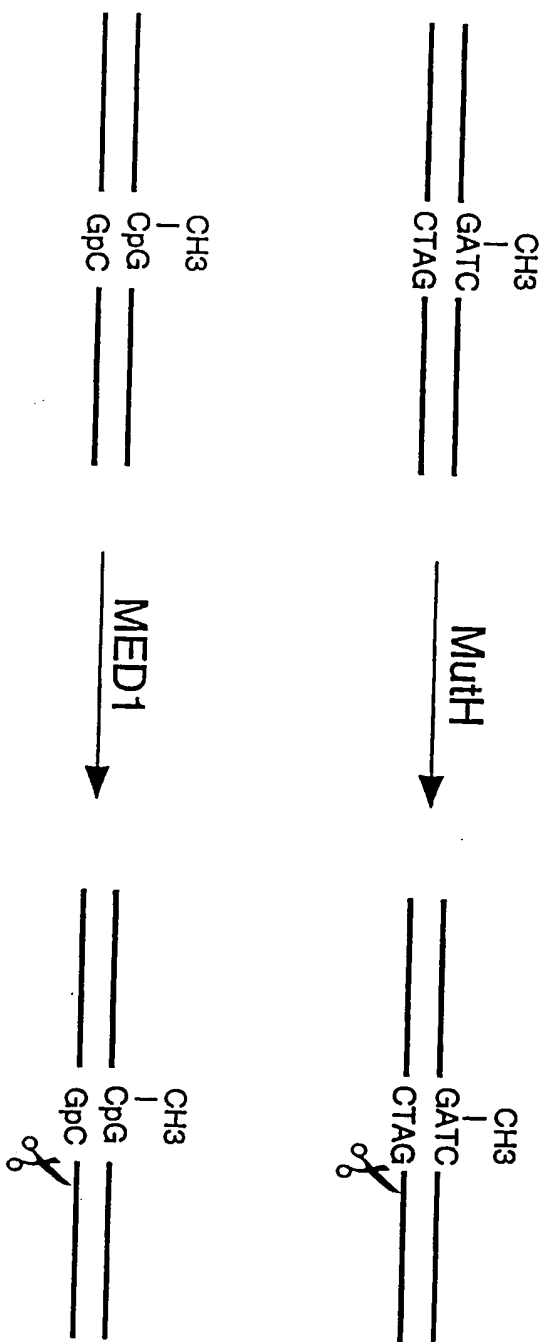


FIGURE 12

Fig. 13A

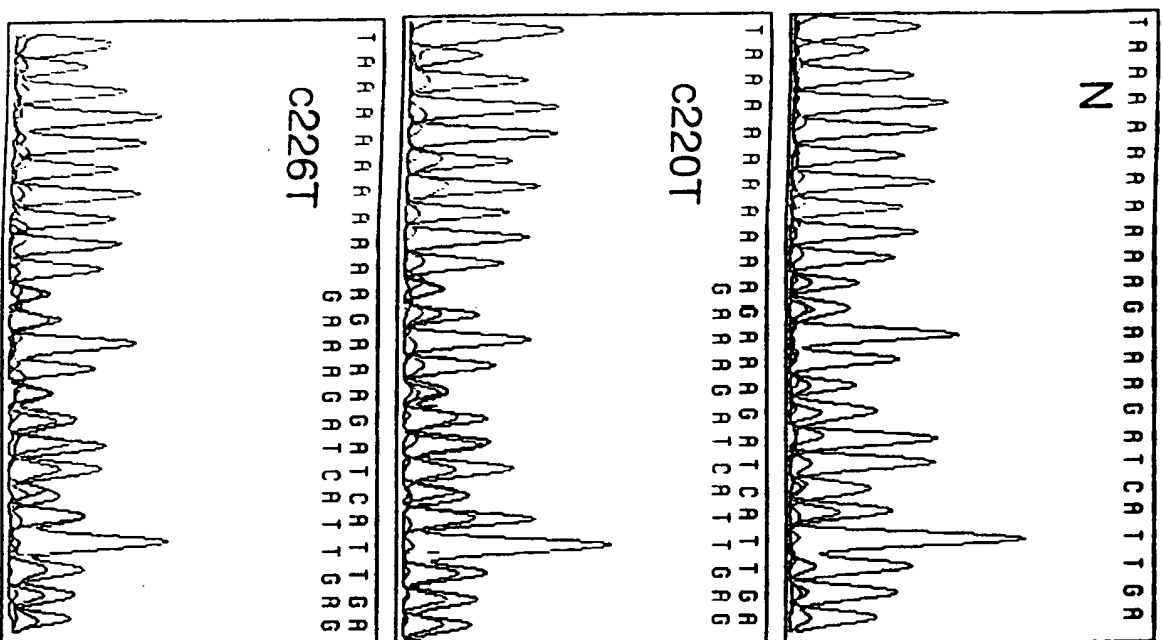


Fig. 13B

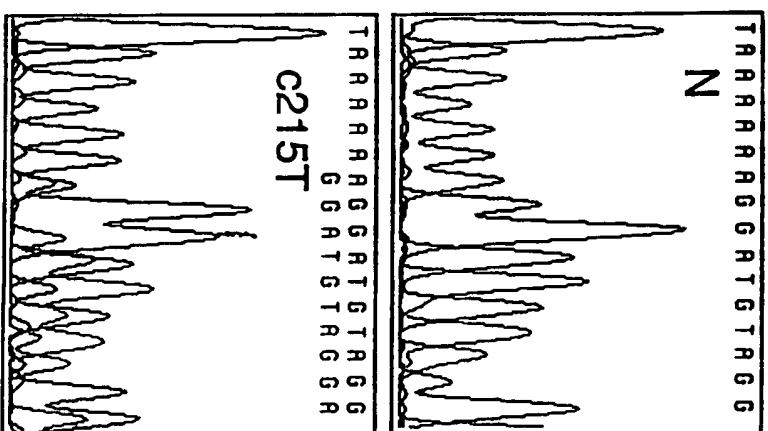
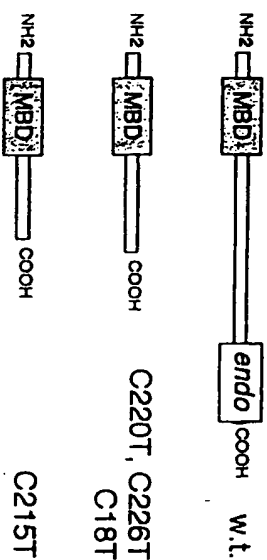


Fig. 13.C



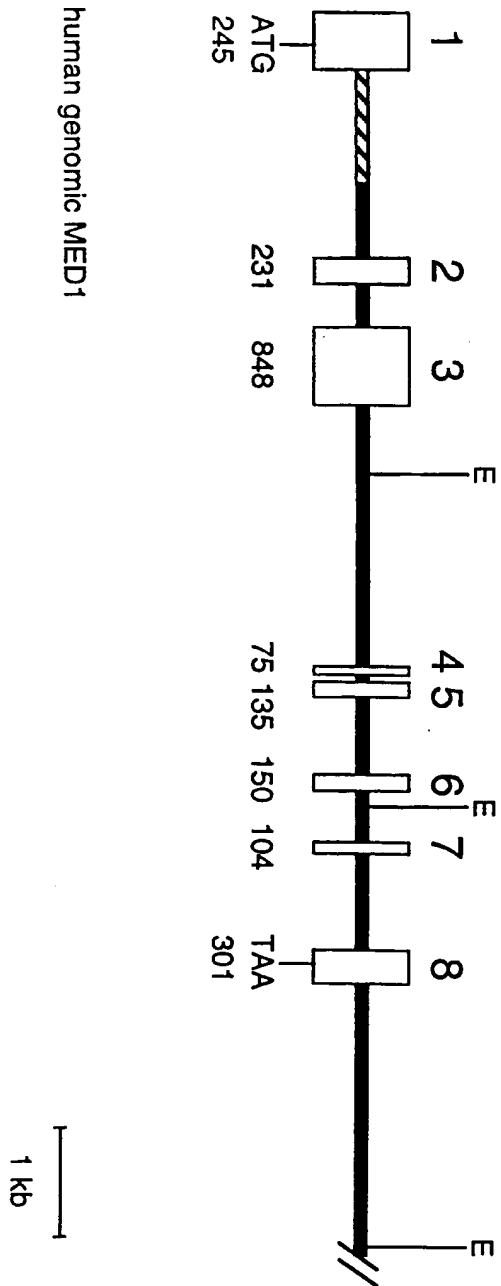


Figure 14

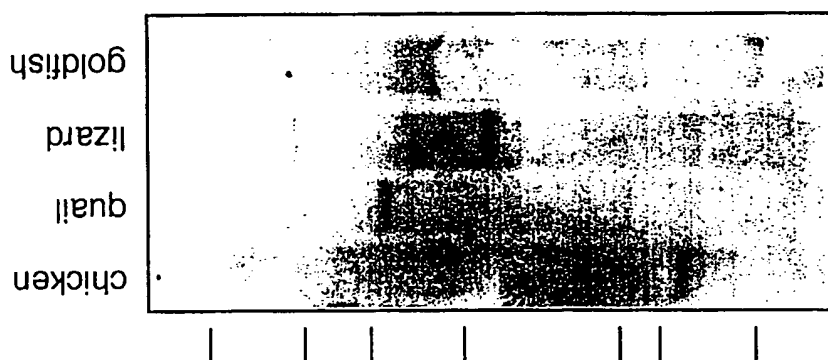


Fig. 15B

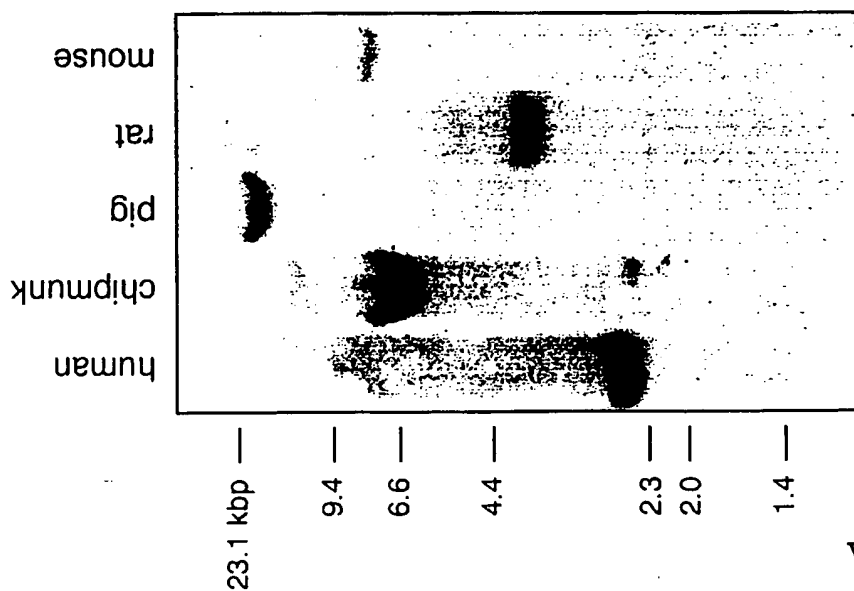


Fig. 15A

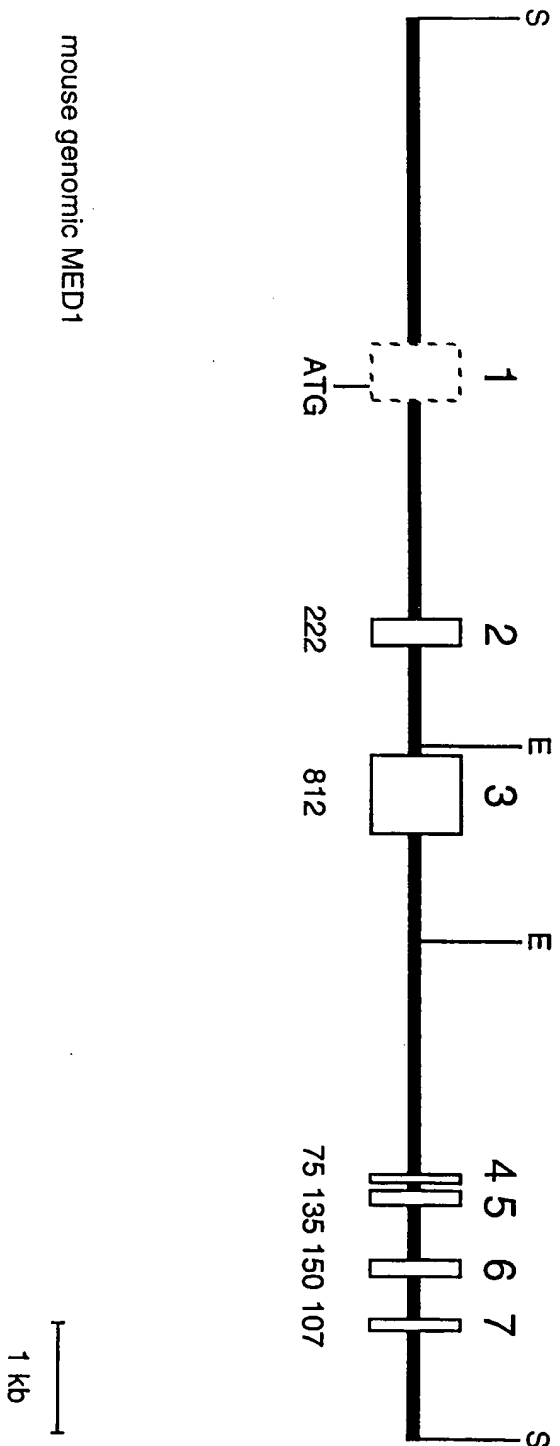


Figure 16

1 CAAGGAAGAT ATTGCTGTTG GACTGGGAGG AGTGGGAGAA GATGGAAAGG
51 ACCTGGTGAT AAGCAGTGAG CGCAGCTCCC TTCTCCAAGA GCCCACTGCT
101 TCTACTCTGT CTAGTACTAC AGCGACAGAA GGCCACAAGC CTGTCCCCTG
151 TGGATGGGAA AGAGTTGTGA AGCAAAGGTT ATCTGGGAAA ACTGCAGGAA
201 AATTTGATGT ATACTTTATC AGCCCACAAG GATTGAAGTT CAGATCAAAA
251 CGTTCACCTG CTAATTATCT TCTCAAAAAT GGGGAGACTT TTCTTAAGCC
301 TGAAGATTTT AATTTTACTG TACTGCCGAA AGGGAGCATC AATCCCGGTT
351 ATAAACACCA AAGTTTGGCA GCTCTGACTT CCCTGCAGCC AAATGAAACT
401 GACGTTTCAA AGCAGAACCT CAAGACACGA AGCAAGTGA AAACAGATGT
451 GTTGCCCTCTG CCCAGTGGTA CTTCAGAGTC GCCAGAAAAGC AGCGGACTGT
501 CTAAGTCTAA CTCGGCTTGC TTGCTATTGA GAGAACATAG GGACATTCAG
551 GATGTTGACT CTGAGAAGAG GAGAAAGTCC AAAAGAAAGG TGACTGTTTT
601 GAAAGGAACT GCAAGTCAGA AAACCAAACA AAAGTGCAGG AAGAGTCTCT
651 TAGAGTCTAC TCAAAGAAAC AGAAAAAGAG CATCTGTGGT TCAGAAGGTG
701 GGTGCTGATC GCGAGCTGGT GCCACAGGAA AGTCAACTCA ACAGAACCCT
751 CTGCCCTGCA GATGCCTGTG CAAGGGAGAC TGTGGCCTG GCTGGGGAAG
801 AAAAATCACC AAGCCCAGGA CTGGATCTTT GTTTCATACA AGTAACTTCT
851 GGCACCACAA ACAAATTCCA TTCAACTGAA GCAGCAGGTG AAGCAAATCG
901 TGAGCAGACT TTTTLAGAAT CAGAGGAAAT CAGATCGAAG GGAGACAGAA
951 AGGGGGAGGC ACATTTGCAT ACTGGTGTTC TACAGGATGG CTCTGAAATG
1001 CCCAGCTGCT CACAAGCCAA GAAACACTTT ACTTCTGAGA CATTTCAAGA
1051 AGACAGCATC CCACGGACAC AAGTAGAAAA AAGGAAAACA AGCCTGTATT
1101 TTTCCAGCAA GTACAACAAA GAAGCTCTTA GCGGAGGAGG ACGCAAATCC
1151 TTCAAGAAAT GGACCCCTCC TCGGTCACCT TTTAATCTTG TTCAAGAAAT
1201 ACTTTTCCAT GACCCATGGA AGCTCCTCAT CGCGACTATA TTTCTCAATC
1251 GGACCTCAGG CAAGATGGCC ATCCCTGTGC TGTGGGAGTT TCTAGAGAAG
1301 TACCCTTCAG CTGAAGTGGC CCGAGCTGCC GACTGGAGGG ACGTGTGCGA

Fig. 17A

1351 GCTTCTCAAG CCTCTTGGTC TCTACGATCT CCGTGCAAAA ACCATTATCA
1401 AGTTCTCAGA TGAATATCTG ACAAAGCAGT GGAGGTATCC GATTGAGCTT
1451 CATGGGATTT GGTAAAATA TGGCAACGAC TCTACCGGAT CTTTGTGTC
1501 AATGAATGGA ACAG

Fig. 17B

mouse MED1 protein (upper sequence) x human MED1 protein
(lower sequence)

```

1  KEDIAVGLGGVGEDGKDLVI..SSERSLLQEPTAST.LSSTTATEGHKP 47
   |||:|. | |||| . :. | ||| . ||||| || . | || |
36 KEDVAMELERVGEDEEQMMIKRSSECNPLLQEPIASAQFGATAGTECRKS 85

48  VPCGWERVVKQRLSGKTAGKFDVYFISPQGLKFRSKRSLANYLLKNGETF 97
   ||||| ||||| |||||:||||| ||||| ||||| ||||| |||||
86  VPCGWERVVKQRLFGKTAGRFDVYFISPQGLKFRSKSSLANYLHKNGETS 135

98  LKPEDFNFTVLPKGSINPGYKHQSLAALTSLOPNETDVSKQNLKTRSKWK 147
   ||||| . |||| | | || | : ||||| | : . . | | : |||| |
136 LKPEDFDFTVLSKRGIKSRYKDCSMAALTSHLQNSNNSNWNLRTRSKCK 185

148 TDVLPPLPSGTSESPSSGLSNSNSACLRLREHRDIQDVDSEKRRKSKRKV 197
   || | || . || | || || | | |||: | : || . | || | || |
186 KDVFMPPSSSELQESRGLSNFTSTHLLLKEDEGVDDVNFRKVRKPKGKV 235

198 TVLKGTSQKTKQKCRKSLESTQNRKRAS..... 228
   |:||| . |||. |||| | . || |
236 TILKGIPIKTKKGRKSCSGFVQSDSKRESVCNKADAESSEPVAQKSQLD 285

.
.
.

229 .....EDSIPRTQVEKRKTSLYFSSSKYNKEALSPPRRKSF 263
   ||. |||||: ||||| ||||| ||||| ||||| ||||| |||||
386 CSPTRKDFTEGKIFQEDTIPRTQIERRKTSLYFSSSKYNKEALSPPRRKAF 435

264 KKWTPPRSPFNLVQEILFHDWPWKLLIATIFLNRTSGKMAIPVLWEFLEY 313
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
436 KKWTPPRSPFNLVQETLFHDWPWKLLIATIFLNRTSGKMAIPVLWKFLEKY 485

314 PSAEVARAADWRDVSELLKPLGLYDLRAKTIKFSDEYLTQWRYPIELH 363
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
486 PSAEVARTADWRDVSELLKPLGLYDLRAKTIVKFSDEYLTQWKYPIELH 535

364 GIWLKYGNDSYRIFCVNEWKQ 384
   || ||||| ||||| |||||
536 GIG.KYGNDSYRIFCVNEWKQ 555

```

Figure 18

A**Exon 2**

ggttttggttttccagCAAGGAAGATATTGCTGTTGGACTGGGAGGAGTG
GGAGAAGATGGAAAGGACCTGGTGATAAGCAGTGAGCGCAGCTCCCTTCT
CCAAGAGCCCACTGCTTCTACTCTGTCTAGTACTACAGCGACAGAAGGCC
ACAAGCCTGTCCCGTGTGGATGGGAAAGAGTTGTGAAGCAAAGGTTATCT
GGGAAAAC TGCAGGAAAATTTGATGTATACTTTATCAGgtaagcatttag
gaaggaaaata

B**Exon 3**

ctttttttttttccttttaagCCCACAAGGATTGAAGTTCAGATCAAAAC
GTTCACTTGCTAATTATCTTCTCAAAAATGGGGAGACTTTTCTTAAGCCT
GAAGATTTTAATTTTACTGTACTGCCGAAAGGGAGCATCAATCCCGGTTA
TAAACACCAAAGTTTGGCAGCTCTGACTTCCCTGCAGCCAAATGAAACTG
ACGTTTCAAAGCAGAACCTCAAGACACGAAGCAAGTGGAAAACAGATGTG
TTGCCTCTGCCCAGTGGTACTTCAGAGTCGCCAGAAAGCAGCGGACTGTCT
TAACTCTAACTCGGCTTGCTTGCTATTGAGAGAACATAGGGACATTCAGG
ATGTTGACTCTGAGAAGAGGAGAAAGTCCAAAAGAAAGGTGACTGTTTTG
AAAGGAACTGCAAGTCAGAAAACCAAACAAAAGTGCAGGAAGAGTCTCTT
AGAGTCTACTCAAAGAAACAGAAAAAGAGCATCTGTGGTTCAGAAGGTGG
GTGCTGATCGCGAGCTGGTGCCACAGGAAAGTCAACTCAACAGAACCCCTC
TGCCCTGCAGATGCCTGTGCAAGGGAGACTGTTGGCCTGGCTGGGGAAGA
AAAATCACCAAGCCCAGGACTGGATCTTTGTTTCATACAAGTAACTTCTG
GCACCACAAACAAATTCCATTCAACTGAAGCAGCAGGTGAAGCAAATCGT
GAGCAGACTTTTTTAGAATCAGAGGAAATCAGATCGAAGGGAGACAGAAA
GGGGGAGGCACATTTGCATACTGGTGTTTTACAGGATGGCTCTGAAATGC
CCAGCTGCTCACAAGCCAAGAAACACTTTACTTCTGAGACATTTCAAGgt
actcagtgcataaaaa

FIGURE 19

C Exon 4

gactataaactaatttttgcttctcagAAGACAGCATCCCACGGACACAAG
TAGAAAAAAGGAAAACAAGCCTGTATTTTCCAGCAAGTACAACAAAGAA
Ggtacccacctttccctaagc

D Exon 5

tatatttntgnagCTCTTAGCCCCCAAGACGCAAATCCTTCAAGAAATG
GACCCCTCCTCGGTCACCTTTTAATCTTGTTCAAGAAATACTTTTCCATG
ACCCATGGAAGCTCCTCATCGCGACTATATTTCTCAATCGGACCTCAGg
ttnggggtcattgncat

E Exon 6

tgtttatgctccccagGCAAGATGGCCATCCCTGTGCTGTGGGAGTTTCT
AGAGAAGTACCCTTCAGCTGAAGTGGCCCGAGCTGCCGACTGGAGGGACG
TGTCGGAGCTTCTCAAGCCTCTTGGTCTCTACGATCTCCGTGCAAAAACC
ATTATCAAGTTCTCAGgtatgtccccagcccag

F Exon 7

tggatgtgtatccctcagATGAATATCTGACAAAGCAGTGGAGGTATCCG
ATTGAGCTTCATGGGATTTGGTTAAAATATGGCAACGACTCTACCGGAT
CTTTTGTGTCAATGAATGGAACAGgtaagcccaccactggggcc

FIGURE 19 cont.

A**Exon 1**

GCGGCGGCGTCTGGGGCGCTTTCGCAACATTCAGACCTCGGTTGCAGCCCGGTGCCGTGAGCTGAA
GAGGTTTCACATCTTACTCCGCCCCACACCCTGGGCGTTGCGGCGCTGGGCTCGTTGCTGCAGCCC
GACCCTGCTCGATGGGCACGACTGGGCTGGAGAGTCTGAGTCTGGGGGACCGCGGAGCTGCCCCCA
CCGTCACCTCTAGTGAGCGCCTAGTCCCAGACCCGCCGAATGACCTCCGgtaagttactgtccccct
tttgggcctcagtttcaccacctgtaaaatgggtatcgggagagtggacagtgtgtgggcctttcta
acctttgacagagggtcggcanaaacctcgaagcccacgggttagttactaggggtctggagccca
ggtgctcttctgtgcatcagc...

B**Exon 2**

...tttggagacaggaaataactcccatagcacaagactggtcacactgactttaatctccc
tcattttaatatggataatctatgtggttcctgcattgtcatggattaaaactgagtaggcagtg
aagataaattttaaataagttaatcacttagactttgtttttccagCAAAGAAGATGTTGCTATGG
AATTGGAAAGAGTGGGAGAAGATGAGGAACAAATGATGATAAAAAGAAGCAGTGAATGTAATCCCT
TGCTACAAGAACCCATCGCTTCTGCTCAGTTTGGTGCTACTGCAGGAACAGAATGCCGTAAGTCTG
TCCCATGTGGATGGGAAAGAGTTGTGAAGCAAAGGTTATTTGGGAAGACAGCAGGAAGATTTGATG
TGTACTTTATCAGgtaagcatataagatggtaaagatagtacagccaaatgattttgtctgggcag
gtagtgggagcatagcaggaatcttagcttctttatatttttaccataaaaccattgcagattc
tattctttcaatgttgctattaattacatcaagtgatttggggaaaattacatacattttgtccct
ccttctgtgaatggttaacgggtaggttgcattttagttatattttataaatttatattgtcataga
ggaaccattttaaaaggccattatcactctttttcatttttaaatgacagagacctatggcaacatt
tggaattaattagaatctgaaatgtggtccagttcttttaaaagtccttctatttactagcagt
aagtttcctttaatatcattttctag(continues into exon 3, see below)

FIGURE 20

C Exon 3

aatctgaaatgtgggtccagttccttttaaaagtcccttctatttactagcagtaagtttccttt
aatatcatttttctagCCCACAAGGACTGAAGTTCAGATCCAAAAGTTCACCTTGCTAATTATCTTCA
CAAAAATGGAGAGACTTCTCTTAAGCCAGAAGATTTTGATTTTACTGTACTTTCTAAAAGGGGTAT
CAAGTCAAGATATAAAGACTGCAGCATGGCAGCCCTGACATCCCATCTACAAAACCAAAGTAACAA
TTCAAACTGGAACCTCAGGACCCGAAGCAAGTGCAAAAAGGATGTGTTTATGCCGCCAAGTAGTAG
TTCAGAGTTGCAGGAGAGCAGAGGACTCTCTAACTTTACTTCCACTCATTTGCTTTTGAAAGAAGA
TGAGGGTGTGATGATGTAACTTCAGAAAGGTTAGAAAGCCCAAAGGAAAGGTGACTATTTTGAA
AGGAATCCCAATTAAGAAAATAAAAAAGGATGTAGGAAGAGCTGTTCAAGGTTTTGTTCAAAGTGA
TAGCAAAAGANAATCTGTGTGTAATAAAGCAGATGCTGAAAGTGAACCTGTTGCACAAAAAAGTCA
GCTTGATAGAACTGTCTGCATTTCTGATGCTGGAGCATGTGGTGAGACCTCAGTGTGAGCAGTGA
AGAAAACNGCCTTGTAATAAAAAAAGAAAGATCATTGAGTTCAGGATCAAATTTTTGTTCTGAACA
AAAAACTTCTGGCATCATAAACAAATTTTGTTTCAGCCAAAGACTCAGAACACAACGAGAAGTATGA
GGATACCTTTTTAGAACTCTGAAGAAATCGGAACAAAAGTAGAAGTTGTGGAAAGGAAAGAACATTT
GCATACTGACATTTTAAACGTGGCTCTGAAATGGACAACAAGTCTCACCAACCAGGAAAGACTT
CACTGgtgagaaaaatatttcaaggtatccagtgctttcagcactattaacattagtgatgagaa
atttatatgctgcatcgtatcgtgccatac

Please note: at the end of exon 3, two alternative splice donor sites are present (see Sequence Variations, page 40 of the application).

D Exon 4 and Exon 5

tagtaccaagttcatgggtcattagttagattaattgggtattttatgtaaagggcttagaatagtg
cctggcatgctttgtaaatagtggtgatattattatttgcatccctcaatattgctttaagcta
aaccatagactccataaagtgtttacttttccttttcagAAGATACCATCCCACGAACACAGATAG
AAAGAAGGAAACAAGCCTGTATTTTCCAGCAAAATATAACAAAGAAGgtatccctttcccaatca
gaacagcaaattctaattccattttgggttttcaattctgatgcactatgtttgtttagCTCTTAG
CCCCCAGGACGTAAAGCCTTTAAGAAATGGACACCTCCTCGGTACCTTTTAATCTCGTTCAAGA
AACACTTTTTCATGATCCATGGAAGCTTCTCATCGCTACTATTTTCTCAATCGGACCTCAGgttt
ggggattattatcatctttgtcttagtagagacagtggtggtagggagaaagcactgaattgag
gcctgggttcaaagtcattttgagtggtgcacctgggatagggcattccccctttcacccttaaac
tcttcacctatgaggaaaatggggg

FIGURE 20 Cont. 1

E

Exon 6

ccagtgttttttgttttttgttttctttaaaaaaaaaaaaaaacctctggatgagatttctatga
gaaactacttgaacgtgaaatcagcccacctggagctcttgtaatcattcagttacttttacnttcc
cagGCAAAATGGCAATACCTGTGCTTTGGAAGTTTCTGGAGAAGTATCCTTCAGCTGAGGTAGCAA
GAACCGCAGACTGGAGAGATGTGTCAGAACTTCTTAAACCTCTTGGTCTCTACGATCTTCGGGCAA
AAACCATTTGTCAAGTTCTCAGgtatttttctatacacccaaaggaaaaacataatacattgtgctt
atthaagagagccacaccttaaaactttaatgtttctcagatactatattaatggaggtttttca
gctcaagcattttaaaaagtccacttttccccaaccacagtctcccactgacctaacaataaat
cttt

F

Exon 7

ctttagaagctgacctgataatgtgggatgttgtattcttcagATGAATACCTGACAAAGCAG
TGGAAGTATCCAATTGAGCTTCATGGGATTGGTAAATATGGCAACGACTCTTACCGAATTTTTTGT
GTCAATGAGTGGGAAGCAGgtgaggctcactcccatccataattcagcacatttggtctctgagg
caaaataagtcaccattatggttaagacnatttattggggatacaaatgctattacagtcacaa
caattgtgttctctggctgcggggaagcgngtggcatgtgggttttgggggttttgatcagtaggcg
ctcccagg

G

Exon 8

tgtgtgagattaccttaatataggtataacttaaaatattcatgaatcccaggaggttaaaggtt
ataacttttaggtatggtatcgtatcgtactgttccccagcaaacatttaaaaagccaatttt
aaaaaatgtatttctgactaagttacatntaaggtctctgcctctgtatcttatgtttcttccagG
TGCACCCTGAAGACCACAAATTAAATAAATATCATGACTGGCTTTCCCAAAATCATGAAAAATTAA
GTTTATCTTAAACTCTGCAGCTTTCAAGCTCATCTGTTATGCATAGCTTTGCACTTCAAAAAAGCT
TAATTAAGTACAACCAACCACCTTTCCAGCCATAGAGATTTTAATTAGCCCAACTAGAAGCCTAGT
GTGTGTGCTTTCTTAATGTGTGTGCCAATGGTGGATCTTTGCTACTGAATGTGTTTGAACATGTTT
TGAGATTTTTTTAAAATAAATTATTATTGACAACA*atccaaaaaaatacggcttttccaatga
tgaaatataatcagaagatgaaaaatagttctaaactatcaataatacaagcaaatttctatca
gccttgctaaagctagggggcccactaaatatttt

Please note: asterisk indicates the poly(A) addition site.

FIGURE 20 Cont. 2

H Complete sequence of the intron between exon 7 and exon 8.

GGAAGCAGgtgaggctcactcccatccataattcagcacatttggtctctgaggcaaaataagtcc
accattatggttaagactatttatttgatacaaatgctattacagtcacaaacaattgtgttcctg
gctgcggggaagcgagtggtcatgtgggttttgggggtttttgatcagtaagcgctcccaagtccaca
aagaccagtcagcggtggcctctgactcatctccagtggtttgtcacctctggccctgttcct
gtcattccctatttgtgtgctatctctaagcctgacgtggttttccctcctgtcaaaagtacaccac
tacaggaaagcaggaagggttgggccttgcaatgtatgcatattgggtttctcttagtggtctcag
actacgtttgtggtgactgggtcctgcttcagccctggtgaatatgccagcctgtggcatgctgg
tggtcatcctggcagctggtgggtggcctggtatgctgcccactcagcttgagactcaccctcatg
cattcagccagtaggtctggccaagcctgaactgaaggacatggtcctatcccagcttcacacaca
gcaatccattgtgacctgagaatccatttaacctctcggtctagaacctccttctggaaagtgagg
tattaatacttgactcaatgttatcgccacccacattctaagtcatggttgagtagtaatttgga
cagtaccttgtaaattgtgtgagattaccttaataataaggtataacttaaaatattcatgaatccc
aggagggttaaagggtataacttttaggtatggtatcgtaatgtactgtccccagcaaacatttaa
aaagccaattttaaaaaatgtatttctgactaagttacattaaggtctctgcctctgtatcttatg
tttcttcagGTGCACCC

FIGURE 20 Cont.3

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US98/15828

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :Please See Extra Sheet.

US CL :Please See Extra Sheet.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : Please See Extra Sheet.

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS, GenBank, MEDLINE, EMBASE, BIOSIS, CAPLUS, SCISEARCH

search terms: human, repair, methyl-cytosine, methyl?, DNA, binding, endonuclease#

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	LEWIS et al. Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA. Cell. 12 June 1992, Vol. 69, pages 905-914, especially page 908.	1-13
Y	Database GenBank, US National Library of Medicine (Bethesda, MD, USA), No. AA011232, HILLIER et al., 'The WashU-Merck EST Project', complete record, 29 July 1996.	11-12

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A" document defining the general state of the art which is not considered to be of particular relevance	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"B" earlier document published on or after the international filing date	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"A" document member of the same patent family
"O" document referring to an oral disclosure, use, exhibition or other means	
"P" document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search

19 OCTOBER 1998

Date of mailing of the international search report

18 NOV 1998

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231

Facsimile No. (703) 305-3230

Authorized officer

SCOTT D. PRIEBE

Telephone No. (703) 308-0196

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US98/15828

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

Please See Extra Sheet.

1. ☒ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US98/15828

A. CLASSIFICATION OF SUBJECT MATTER: IPC (6):

A01K 67/00; A61K 38/46, 39/395; C07H 21/00; C07K 16/18, 16/40; C12N 9/16, 15/12, 15/55, 15/63; C12Q 1/68

A. CLASSIFICATION OF SUBJECT MATTER: US CL :

424/94.6, 130.1, 139.1; 435/6, 7.1, 7.4, 19, 252.3, 254.11, 320.1, 325, 419; 530/358, 387.1, 387.9, 388.1; 536/23.1, 23.5, 24.31; 800/8

B. FIELDS SEARCHED

Minimum documentation searched
Classification System: U.S.

424/94.6, 130.1, 139.1; 435/6, 7.1, 7.4, 19, 252.3, 254.11, 320.1, 325, 419; 530/358, 387.1, 387.9, 388.1; 536/23.1, 23.5, 24.31; 800/8

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION WAS LACKING

This ISA found multiple inventions as follows:

This application contains the following inventions or groups of inventions which are not so linked as to form a single inventive concept under PCT Rule 13.1. In order for all inventions to be searched, the appropriate additional search fees must be paid.

Group I, claim(s) 1-12, 19 parts a, c, d, e, f, 20-22, 24 and 25, drawn to nucleic acids hybridizing to a polynucleotide (SEQ ID NO: 1) encoding human MED1 endonuclease, methods of using same to detect mutant MED1 nucleic acid and kits for carrying out the method.

Group II, claim(s) 13, 17, 27 and 28, drawn to a human endonuclease that binds me-CpG, methods of identifying agents that modulate activity of the protein.

Group III, claim(s) 14-16, 18, 19 parts b and e, 23 and 26, drawn to antibody against human endonuclease that binds me-CpG, method using same to detect endonuclease and kit for the method.

Group IV, claim 29, drawn to a chimeric animal comprising an exogenous MED1 allele.

The inventions listed as Groups I-IV do not relate to a single inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons: The nucleic acids of group I, endonuclease of group II and antibody of group III are structurally, biochemically and functionally different biochemical compounds. They cannot be used for the same purposes, nor substituted for each other. Claim 19 is drawn to different inventions, parts a, c, d, f and part of e are methods of using the nucleic acid, while part b and part of e are methods of using the antibody of group III. The nucleic acid of group I is isolated and must be so in order to use it in the methods of group I as a diagnostic agent. In contrast, the animal of group IV requires incorporation of the nucleic acid and the nucleic acid required in the chimeric animal is not of the same scope as the nucleic acid of group I.